SUPPLEMENTO
AL VOLUME II, SERIE X, DEL
NUOVO CIMENTO
A CURA DELLA SOCIETÀ ITALIANA DI FISICA

1955 2° Semestre N. 4

A Realistic Theory of Mesons.

B. JOUVET

Institut Henri Poincaré - Paris (+)

(ricevuto il 7 Gennaio 1955)

CONTENTS. — § 1. Introduction. – § 2. The theories of nuclear interactions. *a*. The theories F (theory of FERMI, or « pair theories »). *b*. The theories Y (theory of Yukawa, or « mesons theories »). *c*. The theory Y+F. *d*. The realistic theory of elementary particles (theory R). *e*. The general Yukawa theories. – § 3. General Equations and Methods. *a*. Particles and coupling. *b*. Regularization, normalization and renormalization. *c*. The primitive interaction kernels. – § 4. Physical meaning of the interactions: the mesons. *a*. Form of the interaction kernels. *b*. Existence and properties of mesons. *c*. The mesons equations. *d*. The free mesons. *e*. The mesons in beta decay. – § 5. The realistic meson theory and the Yukawa theories. *a*. The Yukawa charged nuclear forces. *b*. The mesons theory of beta decays. – § 6. Generalisation of the realistic theory. *a*. The renormalizations. *b*. The fermion mesons. *c*. The problem of four fields coupling. – § 7. Conclusions.

1 – Introduction.

The explanation of « elementary » particles structure, and therefore the prediction of their characteristics (mass, spin, coupling constants with the various fields) has been the object of much research proceeding mainly from two concepts:

— The concept which we call « realistic », which is an attempt to explain the structure of elementary particles, considered as formed in a definite manner

(+) At present with the Theoretical Study Division of CERN (European Organization for Nuclear Research) at the Institute for Theoretical Physics. University of Copenhagen.

by certain spin $\frac{1}{2}$ particles, supposed more elementary. The main difficulty lies in the way to state, and then to resolve, the problem of the binding of fermions without being given the medium of the boson fields *a priori* [1].

— At the beginning of research in this field, this first concept did not appear to be satisfactory, and more formal methods proceeding from a phenomenological concept of elementary particles were able, if not to explain the relations between the particles, at least to give the form of the equations of the particles supposed free and without structure [2]. Attempts have been made for several years to propose new frames and new principles to represent or deduce either the masses and spins or the coupling of the elementary particles (non-local theories, spin or charge space, mass operators...) [3].

It is the purpose of this paper to present a new form of realistic concept of mesons unlike the previous proposed models [1]. We limit our examples to the pseudo-scalar interactions. We shall show that in order to unify Yukawa theory of nuclear forces and Fermi theory of beta decay (Sect. 2) the choice of a Fermi coupling between the various pairs of fermions (nucleons and leptons) allows us, when using the regularization and a certain summing up of the terms in the S matrix (Sect. 3), to demonstrate the existence of mesons (bosons) whose masses and coupling constants with the various fermion pairs are expressed with the properties of the fermions only (Sect. 4). The construction of a theory of Yukawa type equivalent to the realistic theory allows us to give the rules for the treatment of every term of the S matrix. For certain values of the regularization constants, Yukawa theory is renormalizable (Sect. 5). Various problems concerning the elementary particles are suggested (Sect. 6). This work is the generalization of a more restricted attempt to explain the photon by postulating a Fermi vector coupling between the electron and neutrino pairs (attempt to which we refer here as « V^0 ») [4].

---

[1] L. DE BROGLIE: *Theorie des particules à spin (methode de fusion)* (Paris, 1943); C. N. YANG and E. FERMI: *Phys. Rev.*, **76**, 1739 (1949); G. WENTZEL: *Phys. Rev.*, **79**, 710 (1950); W. HEISENBERG: *Zeits. f. Naturf.*, **5a**, 251, 367 (1950); *Physica*, **19**, 897 (1953); H. YUKAWA: p. 287 in *L. de Broglie, physicien et penseur* (Paris, 1952).

[2] A. PROCA: *Journ. de Phys.*, **7**, 347 (1936); M. FIERZ; *Helv .Phys. Acta*, **12**, 4 (1939); W. PAULI and M. FIERZ: *Proc. Roy. Soc.*, **173**, 211 (1939).

[3] M. BORN: *Rev. Mod. Phys.*, **21**, 463 (1949); B. T. DARLING: *Phys. Rev.*, **80**, 460 (1950); H. YUKAWA: *Phys. Rev.*, **91**, 415, 416 (1953); S. WATANABE: *Phys. Rev.*, **91**, 771 (1953); PAIS: *Physica*, **19**, 869 (1953); A. PROCA: *Journ. de Phys.*, **15**, (1954); O. HARA and T. MARUMORI: *Nuovo Cimento*, **12**, 309 (1954).

[4] B. JOUVET: *Journ. de Math.*, **33**, 201 (1954) or *These de doctorat, Paris* (1954); A summary of V^0 and of the present paper is published in *Compt. Rend. Ac. Sci.*, **237**, 1642 (1953); **238**, 55, 327, 454 (1954); **239**, 1121, 1267 (1954).

2. – The Theories of Nuclear Interactions.

Every theory of the nuclear interactions tries to explain, with the smallest number of concepts, the five interactions (at least):

1) Interactions nucleon-nucleon: (nuclear force).
2) Interactions nucleon-leptons (e, ν, μ): (β-decays).
3) Interactions nucleons-π mesons: (diffusion, absorption or emission).
4) Interactions leptons-π mesons: (π mesons decay).
5) Interactions leptons-leptons: (μ mesons decay).

In quantum theory this problem reduces to the choice of what particles and what couplings between these particles can be considered as existing fundamentally.

We review briefly various solutions.

a) *The theories* F (theory of Fermi, or « pair theories ») ([5]). – The existence of fermions and couplings between the pairs of fermions is enough to provide an explanation of the interactions 1, 2, 5, but this theory does not consider the interactions 3 and 4.

b) *The theories* Y (theories of Yukawa, or « mesons theories »). – The first research on elementary particles led YUKAWA to invent the meson and meson coupling with nucleons, foreseeing interaction 3 and explaining 1. For some years, the lack of experimental precision and the confusion of π-mesons with μ-mesons permitted the hope of explaining the interactions 2 and 5 by coupling meson-fermions, giving also rise to the interactions 4 and 5. This hope no longer exists.

c) *The theory* Y+F. – Since the theories *a*) and *b*) are valid for some interactions and not for others, we are led to postulate that there exist fundamental mesons and fermions coupled with YUKAWA coupling. This postulate explains interactions 3, 4 and 1. Then, to explain the Fermi interactions 2 and 5, we also postulate the existence of a fundamental coupling between the fermion pairs. Thus, the interactions are the result of two types of coupling F or Y, with F larger than Y for the interactions 2 and 5, and Y larger than F for interaction 1. However, to be consistent, it would be necessary to show that the Fermi interactions result effectively from the existence of Fermi coupling. But one does not know how to do this because Fermi coupling is not renormalizable in the sense of Dyson. Then, for lack of a theoretical proof

([5]) See for bibliography: M. TANNENWALD: *Phys. Rev.*, **86**, 332 (1952).

that the corrections to Fermi coupling due to the closed loops of fermions containing unknown constants are negligible, one uses the experimental fact that the Fermi interactions appears to be almost independent of the energy. However this empirical argument seems to us not sufficiently convincing to justify this bicephal theory.

Then a convenient treatment of the Fermi coupling has led us to the possibility of constructing a new type of theory.

d) The realistic theory of elementary particles (theory R). – We postulate that there exist only fermions (nucleons and leptons), and Fermi coupling between the various pairs of fermions. According to the realistic concept of elementary particles, one must be able to deduce from this hypothesis alone not only the form of the interactions 2 and 5 but also the existence of boson particles and the form of their coupling with fermions, and then the interactions 3 and 4, and therefore 1. To show that this program is realisable is the purpose of this paper. It will appear that the theory R is equivalent to a more general Yukawa theory than the usual one.

e) The generalized Yukawa theories. – One can avoid the difficulties concerning the form and the value of the Fermi interaction deduced from the usual Y-theory by postulating the existence of heavy boson mesons coupled more strongly with leptons and more weakly with nucleons than the mesons now known. Such an hypothesis, and the many other which are necessary to explain the high energy nuclear features (hyperons...), often appear arbitrary and thus it is preferable to have a theory of particles. However the price we must pay in the present R-theory is the reject of Dyson renormalization method and the introduction in the theory of a finite number of arbitrary constants in term of which the masses and coupling constants of the boson particles are expressed.

3. – General Equations and Methods.

a) Particles and coupling. – The only particles that we suppose to exist fundamentally are the fermions forming two families, nucleon and leptons. We suppose they have spin $\frac{1}{2}$. We suppose that the only neutral lepton is the neutrino, and denote the leptons by L_0, L_1, $L_2 \ldots L_l$ (L_0 = neutrino).

Let n and p denote the neutron and proton, and N_i the particles of the nucleon family (n, p, hyperons).

In this paper we are interested only in the charged pairs of fermions. All the coupling constants in the charged Fermi coupling are supposed to have the same value, g, of unknown numerical value.

Two pairs which are identical or charge conjugate are supposed uncoupled

The coupled pairs are choosen so that the charge is conserved in elementary processes, and the Lagrangian is Hermitian.

In precise examples, we will utilise the interaction Lagrangian:

$$(3.a1) \qquad \sum_{i=1}^{l} g\{(\overline{\psi}_n\gamma_5\psi_p)(\overline{\psi}_{L_i}\gamma_5\psi_{L_0}) + (\overline{\psi}_p\gamma_5\psi_n)(\overline{\psi}_{L_0}\gamma_5\psi_{L_i})\} + \\ + \sum_{i<j}^{l} g\{(\overline{\psi}_{L_i}\gamma_5\psi_{L_0})(\overline{\psi}_{L_0}\gamma_5\psi_{L_j}) + (\overline{\psi}_{L_0}\gamma_5\psi_{L_i})(\overline{\psi}_{L_j}\gamma_5\psi_{L_0})\}.$$

(Charged pseudoscalar couplings, omission of hyperons).

b) Regularization, normalization and renormalization. – As is well known, the quantum theory of fields, is not usable if one does not add certain principles to allow a non-ambiguous treatment of the infinite expressions. These principles, which often appear as rules for calculus because their meaning is unknown, are the principles of renormalization and otherwise of regularization which is associated with a « normalization » for theories which are not renormalizable. This last occurs in the present case.

We call « regularization », the method which allows us, first to separate in a covariant way the infinities in an ultraviolet divergent expression and, second, to substitute for these infinities some arbitrary constants. This is a mathematical operation, always possible, which associates to every Feynman graph an explicit expression containing a finite number of arbitrary constants. One can imagine that this method supplies a deeper understanding of the infinities problem. The ways to perform the regularization though generally equivalent, are not alike in the concepts that they represent ([6]).

After regularization we must interpret the physical meaning of the resulting expressions, by studying the effect of the arbitrary constants in various concrete physical situations. Two cases can occur:

— The first possibility is that the theory is « Complete » which means that all the arbitrary constants appearing in the theory can be interpreted as corrections to the fundamental constants of the theory, generally contained in the Lagrangian and in the waves normalizations. The existence of invariances of the theory under certain transformations generally permits us to eliminate or to give definite values to some regularization constants (gauge invariances of Maxwell-Dirac theory, or Fermi V^0 coupling ([4])).

— The other possibility is that the theory is « Incomplete » so that the regularization constants represent not only corrections to the constants of the

([6]) W. Pauli and Villars: *Rev. Mod. Phys.*, **13**, 203 (1941); W. Pauli: *Feldquantisierung*, E.T.H. lectures (Zürich, 1950-1951); L. de Broglie: *Compt. Rend. Ac. Sci.*, **229**, 401 (1949); W. Göttinger: *Phys. Rev.*, **89**, 1012 (1953).

theory but also introduce new types of physical concepts not foreseen in the foundations of the theory (Lagrangian); these concepts are new fields or particles and new forms of coupling between the fields (example, the arbitrary meson-meson coupling appearing in the PS theory of mesons).

Once this interpretation has been made an infinite sum of terms depending on an infinite number of arbitrary constants appears in the calculus of a given effect up to an infinite power of the coupling constants. Thus the fundamental constants of the theory are corrected by an infinite number of arbitrary corrections. For this reason, these corrections are not usable. In order to avoid this difficulty the « normalization » principle is defined (7).

— Principle of normalization for complete theories. The arbitrary constants are chosen so that the corrections they give to the fundamental constants of the theory (contained in the Lagrangian), for any order in the development of the S matrix, are null. For example, in meson theory the meson mass and the coupling constant of mesons with nucleons are normalized by postulating that the nucleons closed loops carried by the meson $(B^i(p^2))$ satisfy the condition $(B^i(p^2)/(p^2+\pi^2))_{(p^2+\pi^2)\to 0} \to 0$; therefore:

$$B^i(-\pi^2) = 0\,, \quad \left(\frac{\partial B^i(p^2)}{\partial p^2}\right)_{-\pi^2} = 0\,.$$

This method is used in electrodynamics when one postulates the gauge invariance of the Lagrangian, which is equivalent to postulating the nullity of photon mass in order to eliminate in the S matrix the corrections to the photon mass.

When no more than three wave functions are coupled in the same coupling as in Maxwell-Dirac electromagnetism, the normalization procedure can be expressed mathematically as the conservation of the poles and residues appearing in every degree of approximation of the S matrix.

— Principle of normalization for incomplete theories. One method for the treatment of the physical quantities not included in the Lagrangian consists in completing this Lagrangian by introducing convenient terms containing new constants to be determined by experiments (example, the introduction

(7) We do not at all think that the actual arbitrary regularization and normalization process constitute a satisfactory solution of the infinities problem. But for want of a renormalization method able to treat the fundamental problems of particle structure (mass and coupling), the proposed method is equivalent to Dyson's, when it applies and otherwise, must give satisfactory results if the true, but now unknown, series of corrections to the fundamental constants of the theory are strongly convergent. See, on these problems, the LORENTZ-KAMERLINGH ONNES memorial papers in *Physica*, **19**, (1953).

of λq^4 coupling in meson theory). Another way is to retain the first constant of a new type appearing in the S matrix and to normalize the other constants of that type so that the corrections they add to the first constant are null.

However, the true difficulty of the normalization of theories is the physical definition of the constants. It consists in the choice of the most convenient physical conditions under which a certain interaction has a given value. This difficulty appears more important when there exists a coupling between more than three wave functions, as in the meson-meson coupling of the Yukawa theory and in Fermi couplings; these cases will be discussed in Section **6**.

c) The primitive interaction kernels. – The « pair theories » ([5]) consider only the lowest order interaction terms between two nucleons, via lepton fields. The usual treatment of the Fermi theory takes into account only the first order term in g for the nucleon-lepton interaction. On the contrary, let us define the « primitive interaction kernels », (UV), represented in Feynman graphs as the sum of all the « chains » constitued with any number of closed loops having only two vertices, and connecting a fermion line at a vertex where the pair is U to another (or the same) fermion line at another vertex where the pair is V ([8]).

As examples, with the coupling postulated in (**3**.a), the charged pairs can be:

$$U, V = (np), \quad (L_0, L_1), \quad \ldots, \quad (L_0, L_f).$$

Then we have

(3.c1)

$$\left\{\begin{array}{l} (UU) = \displaystyle\sum_{w \neq u\ u} \; + \sum_{\substack{w \neq u\ u \\ x \neq u \\ x \neq w}} \; + \sum_{\substack{w \neq u\ u \\ x \neq u \\ y \neq x, w}} \; + \ldots \\ \\ (UV) = \displaystyle\sum_{\substack{w \neq u\ u \\ w \neq v}} \; + \sum_{\substack{w \neq v\ u \\ x \neq v \\ w \neq x}} \; + \ldots . \end{array}\right.$$

([8]) S. KAMEFUCHI and H. UMEZAWA: *Prog. Theor. Phys.*, **8**, 579 and **9**, 529 (1953), have shown that such an iteration of loops with two vertices allows us to treat the « second kind » interactions, including the Fermi interactions. However our paper differs from theirs in the method of renormalization, in the demonstration of the possibility of treating completly this interactions, and mainly in the physical interpretation, since these authors have not seen that the Fermi theory is, in the whole, equivalent to the Yukawa theory.

This iteration method, is a generalization of the damping theory of Heitler (*Proc. Camb. Phys. Soc.*, **37**, 291 (1941)) and Wilson (*Proc. Camb. Phys. Soc.*, **37**, 301 (1941)).

The S matrix elements due to these interactions between free particles are

$$(3.c2) \qquad S(UV) = \int \mathrm{d}^4 p J_u^i(p)[\gamma + \gamma^2 \sum_{\substack{w \neq u \\ w \neq v}} K_w^i(p) + \gamma^3 \sum_{\substack{w \neq u \\ x \neq v \\ w \neq x}} K_w^i(p) K_x^i(p) + \ldots] J_v^i(-p) .$$

with $J_u^i(p) J_v^i(-p) = (\bar{a}_\alpha(p_\alpha) O^i a_\beta(p_\beta))(\bar{a}_\gamma(p_\gamma) O^i a_\delta(p_\delta)) \, \delta^4(p_\alpha - p_\beta + p) \cdot$
$\cdot \delta^4(p_\gamma - p_\delta - p)$,

and an analogous formula for $S(UU)$.

Here, the p_λ are the momenta of the four fermions denoted by the indices α, β, γ, δ, paired as $U = (\alpha, \beta)$ and $V = (\gamma, \delta)$. The operator O^i is one of the five Dirac matrix tensors (S, V, T, A, P). The constant $\gamma = ig/2\pi$. The functions $K_w(p)$, associated with the closed loop formed with the pair $W = (\lambda, \mu)$, are

$$(3.c3) \qquad \begin{cases} K_w^i(p) \equiv K_{\lambda,\mu}^i(p) = \int \mathrm{d}^4 x \exp[-ipx] K_{\lambda,\mu}^i(x) , \\ K_{\lambda,\mu}^i(x - x') = \langle P(J_{\lambda\mu}^i(x), J_{\mu\lambda}^i(x')) \rangle_0 ; \qquad J_{\mu,\lambda}^i(x) = \bar{\psi}_\lambda(x) O^i \psi_\mu(x) . \end{cases}$$

Then, one primitive interaction kernel in impulse space is

$$(3.c4) \qquad (UV(p)) = \gamma + \gamma^2 \sum_{\substack{w \neq u \\ u \neq v}} K_w(p) + \gamma^3 \sum_{\substack{w \neq u \\ x \neq v \\ w \neq x}} K_w(p) K_x(p) + \ldots .$$

In this paper it is not necessary to specify the tensor index i.

The algebraically simple form of the kernels results from the fact that the whole impulse is carried in one point of the vertices. More complicated integral equations appear when we take into account the radiative corrections between different loops, that is to say, when we are concerned with closed loops having more than two vertices. This case will be considered in Section **5**.

It is simpler to utilise the equations of which the kernels are solutions rather than the series. One shows easily, from the properties of the Feynman kernels, or by direct substitution, that the primitive interaction kernels satisfy the equations (in impulse space).

$$(3.c5) \qquad \begin{cases} (UU) = \gamma \sum_{w \neq u} (UW) K_w , \\ (UV) = \gamma + \gamma \sum_{w \neq v} (UW) K_w = (VU) = \gamma + \gamma \sum_{w \neq u} (VW) K_w . \end{cases}$$

The solutions are

$$
\left\{
\begin{aligned}
(UU) &= \frac{1}{D}\{\gamma^2 \sum_{w \neq u} K_w + 2\gamma^3 \sum_{\substack{x<w \\ x,w \neq u}} K_x K_w + 3\gamma^4 \sum_{\substack{x<y<w \\ x,y,w \neq u}} K_x K_y K_w + \ldots\} \\
(UV) &= \frac{1}{D}\{\gamma + \gamma^2 \sum_{w \neq u \neq v} K_w + \gamma^3 \sum_{\substack{y<w \\ y,w \neq u \neq v}} K_y K_w + \ldots\} \\
D &= \{1 - \gamma^2 \sum_{y<w} K_y K_w - 2\gamma^3 \sum_{x<y<w} K_x K_y K_w - \ldots\} .
\end{aligned}
\right. \tag{3.c6}
$$

The series appearing in this expressions have a finite number of terms if the number of fermion pairs that we consider is finite.

In the solutions we have omitted arbitrary distributions which appear in the factor $1/D$, (the expression should be $1/D + c\,\delta(D)$ where c is an arbitrary constant). However as we have shown in V^0, this constant is determined by the physical condition that the interactions must be causal, which is true because, the pair exchange is causal. This effect is to shift the poles of the kernel by an arbitrarily small imaginary quantity, as in the definition of the Δ_F functions.

4. – Physical Meaning of the Interactions: the Mesons.

a) *Form of the interaction kernels.* – After regularization, every closed loop having two vertices is a function of the total momentum-energy of the pair and of two arbitrary constants. Without any loss of generality one can suppose that these two constants express some properties of the fermion pair forming the loop. Then for the loop formed with the pair $W = (\lambda, \mu)$, one has

$$
K_w = K_{\lambda,\mu} = K_{\lambda,\mu}(p^2, C^1_{\lambda,\mu}, C^2_{\lambda,\mu}) \tag{4.a1}
$$

(p = impulse, C^1 and C^2 the reg. const.).

For some interval of values $p_a^2 < p^2 < p_b^2$, the function $K_w(p^2)$ can be approximated with a better approximation when the C^i are greater and the interval smaller

$$
K_w(p^2) \approx iF_w \cdot (p^2 + M_w) . \tag{4.a2}
$$

The unidimensional constant F_w and the constant M_w (dimension M^2) are functions of the C^i constants and generally depend weakly on the interval. In this approximation (4.*a*2), the kernel (UU) is the ratio of two polynomials.

the degree of the denominator D being n, and of the numerator $n-1$, where n is the number of different pairs ($w = 1, 2, ..., n$). Then one can write

$$(4.a3) \qquad (UU) \approx i \sum_{l=1}^{n} \frac{A^l_{uu}}{p^2 + \pi_l^2} .$$

The numerators of the kernels (UV) with $u \neq v$ have degree $n-2$, and the denominators are the same. Thus

$$(4.a4) \qquad (UV) \approx i \sum_{l=1}^{n} \frac{A^l_{uv}}{p^2 + \pi_l^2} \qquad \text{with} \qquad \sum_{l=1}^{n} A^l_{uv} \approx 0 .$$

The last equality is not important, because it is due to the approximation (4.a2). In a more general way, let us define independently of the approximation (4.a2)

$$(4.a5) \qquad (UV(p^2)) = \frac{N_{uv}(p^2)}{D(p^2)} \qquad \text{for any } u \text{ and } v = 1, 2, ..., n$$

$$(4.a6) \qquad D(-\pi_l^2) = 0 \; ; \quad A^l_{uv} = \frac{1}{i} \left[(UV(p^2)) \cdot (p^2 + \pi_l^2) \right]_{(p^2+\pi_l^2)\to 0} .$$

The constants π_l^2 and A^l_{uv} defined in this way are calculable functions of fermion masses and C^i-constants of all the pairs, and thus depend only on the properties of the fermions.

b) Existence and properties of mesons. – We limit now the discussion of the preceding equations in postulating that the constants A^l_{uv} and π_l^2 are positive real quantities.

It is then clear that the poles of the kernels (zeros of D) can be interpreted, according to the identity (in the approximation (4.a2)) or to the similarity (in the general case (4.a6)) of the kernels (UV) with the kernels that would give a Yukawa theory postulating the exchange of mesons of mass π_l between the fermions.

But, and this point is essential, to be able to interpret the kernels (UV) as resulting from exchange of mesons, one must be able to define the coupling constant between the π_l meson and the fermion pair u, G^l_u, through the relations

$$(4.b1) \qquad A^l_{uv} = G^l_u \cdot G^l_v .$$

Now this is possible only if there exist the identities

$$(4.b2) \qquad (A^l_{uv})^2 = A^l_{uu} \cdot A^l_{vv} ,$$

between the constants A^l_{uv}.

From (4.*a*5) and (4.*a*6), one sees that this identity is satisfied if

$$(4.b3) \qquad \left[(N_{uv}(p^2))^2 - (N_{uu}(p^2))(N_{vv}(p^2))\right] = 0 \qquad \text{when} \quad (p^2 + \pi_l^2) = 0 \,.$$

We show in the Appendix that, resulting from the hypothesis of the couplings (3.*a*) (namely that the coupling between identical pairs is null, and that the coupling constant between different pairs is the same constant g), this identity is satisfied. The reason is that the quantity in brackets in (4.*b*3) is proportional to the function $D(p^2)$ which, by definition (4.*a*6) is null at every pole, independently of the complex value of the constants A_{uv}^l and π_l^2. This result is also independent of the tensor form of the coupling. It is possible that there exists a more general choice of the relations between the Fermi coupling constants, which allows the possibility of interpreting the kernel with meson concepts, but the essential point we want to underline here is that the equality of the Fermi constants for different pairs of fermions (np-ev, np-$\mu\nu$, ev-$\mu\nu$), which is verified experimentally, implies the existence of bosons through which the Fermi interactions occur.

c) *The mesons equations.* – The field operator of meson must be some combination of the field operators of fermion, since the functional space is determined by the eigen-vectors of the free Hamiltonian which is function of only the fermion fields. Let us write the S matrix for the scattering of two fermions:

$$(4.c1) \qquad S \sim \int \sum_{uv} J_u J_v' (UV(x - x'))\, \mathrm{d}x\, \mathrm{d}x' \,.$$

In the approximation (4.*a*2) one has

$$(4.c2) \qquad (UV(x - x')) \approx \sum_l G_u^l \cdot G_v^l \cdot Y_l(x - x') \,,$$

where the Y_l functions are the Yukawa potentials associated with mesons π_l. Then

$$(4.c3) \qquad S \approx \int \sum_u J_u \sum_l G_u^l \, \Pi_l(x)\, \mathrm{d}x \,,$$

with the definition of the meson field operator $\Pi_l(x)$

$$(4.c4) \qquad \Pi_l(x) = \int Y_l(x - x') \Big(\sum_v G_y^l J_v(x')\Big)\, \mathrm{d}x' \,.$$

We have

$$(4.c5) \qquad (\Box - \pi_l^2)\, \Pi_l(x) = \sum_v G_v^l \cdot J_v(x) \,.$$

The exact form of the meson field operator, independently of the approximation (4.a2) is given in Sect. 5.

d) *The free mesons.* – In the S matrix formalism we cannot represent the free mesons (present in $\pm$ infinite time), and the mesons appear always as intermediate states. Physically such a situation seems satisfactory since we do not know stable mesons. However, as Prof. H. UMEZAWA has pointed out to us, if the poles are real and positive quantities ($\pi^2 > 0$), the addition of an arbitrarily small imaginary quantity required to define the Δ_F functions implies the existence of outgoing and ingoing free waves which are not null at plus or minus infinity, which is inconsistent with the S matrix formalism. The way to switch off the meson waves at infinity is to introduce a finite imaginary part in the mass definition giving rise to an exponential decrease of the wave amplitude, associated with an increasing of the free pairs of fermions. Such a situation appears in this theory if, and only if, the masses π_i^2 are larger than the energy necessary to allow a real creation of some fermion pair; indeed the $K_{\alpha\beta}(p^2)$ functions present a finite imaginary part when $-p^2 > (m_\alpha + m_\beta)^2$, and then the D function is a complex one (9). If further information on this problem makes this interpretation correct, then to avoid in V^0 theory (4) the presence of photons at infinity, it would be necessary to have a finite imaginary part of the mass. According to this theory, this would mean that the neutrino mass is not absolutely null. However, this has no effect on the gauge invariance, which results only from the vectorial form of the coupling and not from the nullity of the neutrino or photon mass.

Another possible way to attack the problem of particle structure is to utilise some non-adiabatic method (TAMM-DANCOFF or BETHE-SALPETER). However if it is possible to obtain the masses of the mesons in this way it is not possible to deduce also the meson-fermion coupling constants. Furthermore, we have difficulties in this method, in interpreting the part of the mass operator having complex value.

This difficulty seems connected with the fact that one cannot consider the mesons as formed by fermions « bound » together in the usual sense of this word, which implies the notion of persistence of every constituent. In our concept, the meson is successively in every possible state of virtual pairs of fermions, and it appears as some « fusion » state due to the fluctuation of the pairs. The same difference exists in the prediction of the mass of «bound » or « fused » particles. The mass of the particle resulting from the binding of two particles is, for ordinary couplings, not very far from the sum of the masses of the constituents. In fusion, on the contrary, one has no idea of

(9) The presence of complex poles in fermion self energy contributions, and mathematical difficulties thereby arising have been indicated by G. FELDMAN: *Proc. Roy. Soc.*, A **223**, 112 (1954).

the resulting mass, and the intuitive idea that the mass is some mean value of the masses of pairs is wrong, as can be shown by the example of the photon for which the mass is independent of the masses of electron, proton and other charged pairs treated on the same footing as the neutrino pair.

e) *The mesons in beta decay.* – As a result of the preceding discussions (Sect. 2, and 2.*b*), the beta decay necessarily occurs through the meson fields. Precise experiments would be able to detect the mesonic influences. They can be a modification to the lifetime due to interaction of the virtual mesons with the meson, nucleons, coulomb field of nucleus, or a possible bremsstrahlung of the mesons before the decay. We note that the actually known mesons are not sufficient to explain the experimental result (2.*b*).

5. – The Realistic Meson Theory and the Yukawa Theories.

We treat now precise examples, using the Lagrangian (3.*a*1).

a) *The Yukawa charged nuclear forces.* – Let us suppose, as a first approximation, that the regularization constants of the lepton loops are infinite

$$K_{L_i} = \text{infinite} \tag{5.a1}$$

and that the only other pair is the charged nucleon pair (np). Then the primitive interaction kernels are

$$(L_i L_j) = 0\,, \qquad (N L_i) = 0\,, \qquad (NN) = -\frac{l\gamma}{1 + l\gamma K_N}\,. \tag{5.a2}$$

where we denote $L_i = (L_i, L_0)$, $i = 1, 2, ..., l$; and $N = (np)$. In this approximation (5.*a*1) there exists interaction between the nucleons and not between the leptons nor between leptons and nucleons. We see that D generally has one pole. Thus the theory may be equivalent to the usual Yukawa theory of charged nuclear force.

α) *One method of treating the interactions.* – In a general way, we write the K_N function in the form

$$K_N(p^2) = a + b\cdot(p^2 + \pi^2) + (p^2 + \pi^2)\cdot R(p^2)\,, \qquad R(-\pi^2) = 0 \tag{5.a3}$$

the constants a and b being functions of the $C_N^{(1)}$ and $C_N^{(2)}$ regularization constants of the nucleon loop (the exact form of these functions depend upon the method of regularization, and it is not necessary to make this dependence precise here). In this formula we supposed that the meson mass is known. This is without justification since it is a function of the $C_N^{(i)}$. Let us, however, choose the $C_N^{(i)}$

constants in such a way that we have

$$a = -1/l\gamma\,, \qquad b = i/(G^2)\,. \tag{5.a4}$$

Then

$$(NN) = iG^2 \frac{1}{(p^2+\pi^2)(1 - iG^2 R(p^2))}\,. \tag{5.a5}$$

We recognize here the interaction kernel produced by the exchange of a meson, the propagator of which is modified in the first order in G^2 by the vacuum polarization in a Yukawa theory in which the meson is coupled with the nucleon pair by the constant $2\pi G$. It is the Dyson function Δ'_F in which we have omitted the terms proportional to degrees in G higher than the second. We denote this function as Δ'^1_F.

Then, in the same way as for the photon (Chap. 9-A in V[0]) one can show that, from the equality $(NN) = iG^2\Delta'^1_F$, (see equat. (5.a5)), the value of a S matrix element in R theory, represented by the sum of every graph formed with certain open lines of nucleons and certain closed loops of nucleons having more than two vertices but connected from one vertex to another by « chains » having any number of closed loops with two vertices, of lepton and nucleon type, is equal to the value of a corresponding S matrix element, in the Yukawa theory, formed by the same open lines of nucleons and the same nucleon closed loops having more than two vertices, but connected from vertex to vertex by first order corrected meson lines having the value Δ'^1_F. It results from this theorem that, in the (5.a1) approximation the R theory is identical to the Yukawa theory ignoring the lepton fields.

β) *A second way of treating the interactions.* – Although the first method is able to lead to the usual Yukawa theory, we think that this treatment is not physically correct because it implies the *a priori* knowledge of the meson mass and it chooses the constants in order to obtain exactly the usual Yukawa theory. Rather, let us write

$$(NN) = iA/U(p^2, g, C^1_N, C^2_N) \tag{5.a6}$$

and define the experimental constants (A and π^2) by

$$\left\{\begin{array}{l} U(-\pi^2, g, C^1_N, C^2_N) = 0 \rightarrow \pi^2 = \pi^2(g, C^1_N, C^1_N)\,, \\ \left[\dfrac{\mathrm{d}U}{\mathrm{d}p^2}\right]_{(p^2+\pi^2=0)} = 1 \rightarrow A = A\,(g, C^1_N, C^2_N)\,. \end{array}\right. \tag{5.a7}$$

Conversely, the regularization constants C_N^i are

$$C_N^i = C_N^i(\pi, G, g)\,. \tag{5.a8}$$

To be able to generalise the equivalence theorem between the R-theory and some Yukawa theory Y_R, as we have done in the preceding method, let us define the free meson propagator $(1\ X(p^2)$ by the condition that the interaction kernel between two nucleons, due to the exchange of this meson, the propagation of which is corrected to the first order in G^2 by the vacuum polarization terms conveniently normalized, be identical with the primitive interaction kernel (NN) of the realistic theory

$$\left\{iG^2\,\frac{1}{X(p^2)\cdot[1-iG^2\cdot R(p^2)\cdot(p^2+\pi^2)/X(p^2)]}\right\} = \tag{5.a9}$$

$$= -\left\{(NN) - iG^2\cdot\frac{1}{U(p^2,\ g,\ C_N^1,\ C_N^2)}\right\}.$$

From this there results

$$X(p^2) = U(p^2) + iG^2R(p^2)\cdot(p^2+\pi^2)\,. \tag{5.a10}$$

The function $X(p^2)$ is generally not equal to $(p^2+\pi^2)$, because

$$\left|\frac{\mathrm{d}^2(X(p^2))}{(\mathrm{d}p^2)^2}\right|_{-\pi^2} = \left[\frac{\mathrm{d}^2U}{(\mathrm{d}p^2)^2} + 2iG^2\frac{\mathrm{d}R}{\mathrm{d}p^2}\right]_{-\pi^2}, \tag{5.a11}$$

is different from zero, except for particular values of the C_N^i constants (that of the preceding case) (*). Thus

$$X(p^2) = (p^2+\pi^2)(1 + Z(p^2, \pi^2))\,, \qquad Z(-\pi^2, \pi^2) = 0\,. \tag{5.a12}$$

and the Hamiltonian of the free meson is generally not a linear function of π^2. Then the meson mass renormalization method is not suitable and we are obliged to use the normalization. The charge renormalization can still be made with Dyson formalism. Also there results from the condition (5.a9) that the Yukawa theory Y_R, in which the free meson equation is defined by

$$X(p^2)H(p) = 0\,, \tag{5.a13}$$

where $X(p^2)$ is given by (5.a10), is identical to the realistic theory. From this identity, one obtains immediately a simple way to treat every effect deducible

(*) See correction in the Note added at the end of the paper.

from the S matrix. The R-theory is then renormalizable, or normalizable, according to the values that we give to the first regularization constants C_N^i. If it is renormalizable, it is equivalent (in this approximation (5.*a*2)) to the usual Y-theory, and if not one sees that the difference from the usual Y-theory lies in the non-linearity of the meson equation, mostly affecting the results in the virtual meson processes.

b) *The mesons theory of beta decays.* – There exist nucleon-leptons and leptons-leptons interactions when the function K_{L_i} are finite. A new approximation of the R-theory is

$$(5.b1) \qquad |\gamma K_{L_i}| \gg 1 \text{ and finite}, \qquad \gamma K_{L_i} \approx \text{independent of } p^2 \text{ for } |p^2| < M_N^2 .$$

We shall show that this approximation gives rise to the π-meson theory of PS beta decay.

In this approximation the D function has only one pole with low mass. We interpret this pole as the π-meson. Taking for K_N the same function as in (5.*a*), one sees that

$$(5.b2) \qquad \left(\frac{G_{L_i}}{G_{L_j}}\right)^2 = \frac{N_{L_i, L_i}}{N_{L_j, L_j}} \approx \left(\frac{K_{L_j}}{K_{L_i}}\right)^2,$$

G_{L_i} = coupling constant between π-meson and $L_i L_0$ pair.

One obtains the experimental G constants for coupling of π-meson with $(\mu\nu)$ and $(e\nu)$ pairs in choosing

$$(5.b3) \qquad \gamma K_{L_1} \sim 5 \cdot 10^9 , \qquad \gamma K_{L_2} \sim 1.5 \cdot 10^8 .$$

Evidently, in such an approximation, the Fermi interaction resulting from the exchange of the π-meson alone is too weak, mainly the Fermi interaction between the lepton pairs. This results from neglecting the p^2 dependance of K_{L_i} functions, which gives rise to new and heavy bosons more strongly coupled with leptons and less with nucleons.

However, although this is a very crude approximation, one sees appearing a law for the cut-off which seems to have a profound significance: For the pairs of a same family, the magnitude of the function K_w is larger when the mass of the pair is smaller. One can derive the behaviour of this law from a theorem of Heisenberg ([10]) which implies the impossibility of having a local, causal and covariant interaction between null mass particles. Then if we postulate a Fermi coupling between two pairs of fermions, one having a null mass, the interaction between two null mass fermions through the virtual pairs

([10]) W. Heisenberg: in *L. de Broglie, physicien et penseur* (Paris, 1953), p. 283.

(which is causal, local and covariant) must be null. This is the case only when the null mass closed loops have an infinite value. Then, the regularization constants must be choosen so that they tend to infinity when the pair mass tends to zero. This theorem applies quite generally and permits us

(i) to explain the photon structure,
(ii) to forbid the two neutrino decay of π^0-meson,
(iii) to show the impossibility of a local interaction between photons.

6. – Generalization of the Realistic Theory.

a) *The renormalization.* – From the equivalence theorem, between the R-theory, and the Y_R-theory we are led to normalize or renormalize according to the general scheme of sect. 3. Physically we preserve the properties of mass and coupling constants of the new particles which result in the summing of a certain class of graphs. Mathematicaly, this preservation is equivalent to the hypothesis that the poles and residues occurring in this sum of graphs must remain the same when we take into account the other graphs of the S matrix. However, it is not clear in this method why the known beta interaction have almost the same value which appears rather independent of the exchanged energy. This point can be cleared up only by a better knowledge of all the mesons through which this interaction occurs. Another way to renormalize, given by KAMEFUCHI-UMEZAWA (8), starts, on the contrary, preserving the Fermi constant so that at zero energy the Fermi interaction is equal to the Fermi constant. However one does not understand the physical meaning of the terms introduced in the Lagrangian for renormalization. Furthermore, the authors are not concerned with the poles of the kernels, and we do not know if the corresponding masses are also renormalized *de facto*. In R-theory we obtain an equality of the Fermi interactions for low energy, by postulating that every $\gamma K_\mu(p^2)$ function has the same value when p^2 is null.

In our method we dispose of $2L$ arbitrary constants if the g constant is given for every coupled pair (L = number of unlike pairs); then there are (at least) L mesons each coupled with all the L pairs so that we have $L(L+1)$ experimental constants to determine the value of the $2L$ arbitrary constants. Therefore, there must exist relations between the various mesons and one can hope that a better knowledge of the light mesons will help us in research on the heavy mesons. As the R-theory gives a relation between the number of fermion pairs and the number of bosons, so the Y-theory, when it postulates many bosons of the same tensorial nature, must also postulate a corresponding number of fermion pairs to be renormalizable (*).

(*) These points are more precisely discussed in the Note at the end of the paper.

b) *The fermion mesons.* – In the same way as in the preceding sections, one can also build a realistic theory for the complex structure of fermion mesons formed by a definite fusion of certain bosons and certain fermions supposed more elementary and given *a priori*. An example of such a meson appears in the problem of π-nucleon scattering, and is due to the pole produced by the iteration of nucleon self-energy graphs. However as in sect. 4.*d*, this pole can be interpreted as a new meson (here an hyperon, Y) only in the S matrix formalism which allows to define the coupling form $(\psi_Y O^5 \psi_N)\varphi_\pi \cdot G^5_{Y,N}$.

The general argument that we have given about the bosons applies also in this case so that the mass of these new fermions must be complex. This can happen either if the mass of the fermion meson is larger than the mass of its constituents or if the constituents are also meson (complex) particles. The problem of renormalization in this case has already been resolved by the Dyson method. Apparently, the only difficult problem lies in the choice of the most important fundamental particles and in a satisfactory treatment of the strong coupling pion-nucleon. Thus one can hope that if this rather long calculations can be carried out one will be able to construct by degrees all the mesons by postulating some fundamental Fermi coupling between a very restricted number of « most elementary » particles (electron, neutrino, proton and perhaps neutron) (*).

c) *The problem of four fields coupling* ([11]).

α) *The Fermi couplings.* – In the main, we have shown that the Fermi coupling is equivalent to a sum of Yukawa-Lorentz type coupling between the given fermions and new boson particles (mesons) always unstable. There appears also an additional coupling term between the meson wave functions (except in V^0 case) as the $\lambda\varphi^4$ usual term of Y-theory, introduced by MATTHEWS.

Let us now start from the same Lagrangian, and make a Fierz transformation ([12]); then the coupled lepton and nucleon pairs will be reorganized in mixed lepton-nucleon pairs coupled together with different coupling constants and different tensor forms, giving a new form for the Lagrangian. The new meson theory that one can deduce from this new Lagrangian must be identical to the preceding and then there must exist relations between the new and old types of mesons. For the coherence of the S matrix formalism, these new mesons must have masses larger than the proton mass in order to prevent the presence of such meson at infinity by the imaginary part of the mass.

(*) See Note added at the and of the paper.

([11]) We are indebted to H. UMEZAWA for informing us that S. KAMEFUCHI has published a paper on this subject (*Prog. Theor. Phys.*, **11**, 273 (1954)). This review is not arrived in Paris, and it seems from *Phys. Abstract*, 10789 (1954), that this author has not reached the physical conclusion given here.

([12]) FIERZ: *Zeits. f. Phys.*, **104**, 553 (1937).

The hypothesis of such mesons has been introduced in the papers of TANIKAWA and TANAKA ([13]). These mesons decay in nucleon plus lepton, (n, p, Y, ...) → (e, ν, μ ...), according to the value of their mass. These authors showed also that only scalar and pseudoscalar mesons of this type are sufficient to deduce the Yang-Tiomno interactions. Then, according to our theory, one can show that the S and PS Fermi couplings between the mixed pairs (nucleon and lepton) quite suffice to explain the known charged interactions. The R-theory allows us to investigate the existence and the properties of the neutral Fermi couplings. Among the neutral Fermi couplings that experiment indicates, certainly, the Vectorial and PS exist which give rise to the photon and π^0-meson([1]). The existence of θ-mesons which may be scalar can be explained in two different ways. The first way is to derive their existence from S-Fermi couplings (charged and neutral) between the fermion pairs. In this model the resulting meson will be called « primitive » like the π-mesons.

A second way to construct S mesons exists in the PS mesons theory:

β). *The Matthews term in PS meson theory.* – Indeed, for the $\lambda\varphi^4$ term, one can try to apply the same treatment that we have done for the Fermi coupling. However, a new difficulty arises from the fact that one can iterate the meson closed loops in every vertex in many different ways, because one can group in pairs the four mesons lines of a vertex in three different ways. The same difficulty occurs also in Fermi interaction when instead of the Lagrangian that we have considered one uses the sum of the three different Lagrangians obtained by Fierz transformations. However, due to the apparent success of the iterative method in Fermi interaction, we shall suppose that the combination problem of iteration of graphs also has a solution in the meson-meson interaction. Then, as in the Fermi coupling treatment, there appear two arbitrary constants, the λ constant of coupling, and the cut-off constant in meson closed loops with two vertices c_π. Pole and its residue produced by the iteration method are interpreted as due to a new meson having a certain mass and a coupling constant with PS meson pair. This meson is scalar and its structure is different from the S meson constructed with the S-Fermi coupling. We call it a « secondary » meson, and to recall its PS-Fermi coupling origin, we note it S_{PS}.

As an application, if the λ constant is large, there is no reason to neglect it in the calculation of the nuclear potentiel. Its action is important as soon as there are more than two mesons in flight; but then the most important quantity is the pion-S_{PS} coupling constant. In the same way this constant must be introduced in the pion-nucleons scattering, before the usual iteration of nonadiabatic methods.

([13]) TANIKAWA: *Prog. Theor. Phys.* 3 338 (1948); TANAKA: *Prog. Theor. Phys.* 9, 169 (1953).

As a conclusion of this rather imprecise study we want to point out the fact that it seems generally possible to substitute for a more than three field coupling an equivalent description, using additional field quantities and new coupling containing no more than three fields.

7. – Conclusion.

To summarize this paper, we have shown that, with the use of regularization and a certain way of ordering and summing the matrix elements, the Fermi coupling can be completely treated. The main result is the interpretation of the method in terms of new particle concepts. Then we derive from the existence of Fermi coupling between fermion pairs the properties of boson mesons through which the Fermi interactions must occur. The extension of this method and of the physical interpretation can be made, and one reaches the conclusion that every meson, boson or fermion, can be built from the smallest number of fundamental particles (electron, neutrino, proton, and perhaps neutron), requiring no essentially new principle, or new methods. However, the use of regularization, introducing in the theory arbitrary constants to be determined by experiment need to be improved. It is in research on the law of the regularization constants (whose qualitative form is given) and in its further interpretation that one can make progress. Most of the problems underlined in Sect. 6 require further works. We shall treat in our next publications more precise examples of the applications of this theory:

— The PS theory (structure of mesons, bosons and fermions, charge symmetry and independence of nuclear interactions).

— The electromagnetic properties of mesons (charge and photo-decays).

— The V, PV, T interactions (allowed meson coupling forms, gauge invariances).

* * *

The author is indebted to Prof. H. UMEZAWA for pointing out his prior results ([8]) on Fermi coupling treatment, and for his valuable discussions on this paper. Thanks are also due Prof. L. DE BROGLIE and A. PROCA for a critical reading of the French manuscript. We should like to thank ELIZABETH SCOTT for helping in translation.

Appendix

Demonstration of formula (4.*b*3).

The kernel equations are (3.*c*5)

a) $$(UU) = \gamma \sum_{v \neq u} (UV)\cdot K_v \,,$$

b) $$(UV) = \gamma + \gamma \sum_{w \neq v} (UW)\cdot K_w = (VU) = \gamma + \gamma \sum_{w \neq u} (VW)\cdot K_w \,,$$

then

$$(UU) = \gamma\Big[(UV)\cdot K_v + \sum_{w \neq v} (UW)\cdot K_w - (UU)\cdot K_u\Big]$$

$$= \gamma(UV)\cdot K_v + \big((UV) - \gamma\big) - \gamma(UU)\cdot K_u \,, \qquad \text{from } b)$$

or

$$(UV)\cdot(1 + \gamma K_v) = \gamma + (UU)\cdot(1 + \gamma K_u)$$

and by permutation of u and v

$$(VU)\cdot(1 + \gamma K_u) = \gamma + (VV)\cdot(1 + \gamma K_v) \,.$$

The product of these two equations is

$$(UV)^2\cdot(1 + \gamma K_u)\cdot(1 + \gamma K_v) = (UU)\cdot(VV)\cdot(1 + \gamma K_u)\cdot(1 + \gamma K_v) + \\ + \gamma^2 + \gamma\big[(UU)\cdot(1 + \gamma K_u) + (VV)\cdot(1 + \gamma K_v)\big]$$

and using *a*)

$$\big[(UV)^2 - (UU)\cdot(VV)\big]\cdot(1 + \gamma K_u)\cdot(1 + \gamma K_v) = \gamma^2 + \gamma^2\cdot\Big[\sum_w \big((UW) + (VW)\big)\cdot K_w\Big]$$

writing

$$(UV) = \frac{N_{uv}}{D} \quad \text{for every } u \text{ and } v;$$

$$(N_{uv}^2 - N_{uu}N_{vv}) = \gamma^2\cdot\left\{\frac{D + \sum_w (N_{uw} + N_{vw})\, K_w}{(1+\gamma K_u)(1+\gamma K_v)}\right\}\cdot D \,.$$

q.e.d.

Note added in proof.

I meant to have a paper (R) published in *Nuovo Cimento* under the title of « A Realistic Theory of Mesons ». Owing to an unexpected delay, two other papers presented at the Pisa conference, vid.,

(E): Les Couplages de Fermi et la Théorie des Bosons

(F): L'exclusion des Fantômes

will be issued before the first paper (R). I therefore endeavoured to write both of them so that they might be thoroughly understood without precise reference to (R).

I am going to deal with the alterations and generalizations that have occured

1. – On the mesons equations.

Corrections and additions to the Sect. 5*a and* 6*a of* (R). The paper (E) is a general mathematical demonstration of the Equivalence of the R-theory to a certain Yukawa theory, when no real meson is present. This theorem has been very shortly indicated at the end of sect. $5a\alpha$ of (R). The application of (E) permits to show that the mesons equations are linear functions of the Dalembertian, and to clarify their meaning and the methods of renormalization.

Aa) One pair; first order normalization.

Let us first correct the wrong result of non-linearity of mesons equations (R. $5a\beta$) which comes from the fact that we did not utilise the explicite form of the (NN) kernel. The equation (R. 5a9), (which is the condition $(UV_F) = (\bar{U}V_F^1)$ of (E)), has the form

$$\text{(N1.1)} \qquad \left\{(NN) = \frac{\gamma'}{1-\gamma' K_N}\right\} = \left\{\frac{iG^2}{X(p^2)(1-i^2 G K_N^R/X(p^2))} = iG^2 \Delta_F'^1\right\}.$$

We use here the notation γ' for $(-l\gamma)$, and K_N^R for the normalized (or renormalized) expression for the closed loop B_N^2 such that $K_N^R(-\pi^2) = 0$ and $((\partial/\partial p^2) K_N^R(p^2))_{-\pi^2} = 0$. Please remark that the (NN) kernel is identical to the kernel which would result from a direct coupling between nucleon pairs with the coupling constant $g' = -lg$. Expressing K_N^R in term of the regularized expression for closed loop B_N^2, the function K_N, it follows

$$\text{(N1.2)} \qquad K_N^R(p^2) = K_N(p^2) - K_N(-\pi^2) - K_N'(-\pi^2)\cdot(p^2+\pi^2)$$

$$K_N'(-\pi^2) = \left(\frac{\partial}{\partial p^2} K_N(p^2)\right)_{-\pi^2}.$$

Then, from (N1.1) and (N1.2) one obtains

$$\text{(N1.3)} \qquad \{iG^2 \Delta_F = iG^2/X(p^2)\} = \frac{\gamma'}{1-\gamma'(K_N - K_N^R)},$$

$$\text{(N1.4)} \qquad \frac{X(p^2)}{iG^2} = \frac{1}{\gamma'} + (K_N^R(p^2) - K_N(p^2)) =$$

$$= \frac{1}{\gamma'} - K_N(-\pi^2) - K_N'(-\pi^2)\cdot(p^2+\pi^2),$$

$X(p^2)$ is a linear function of p^2; the constants $K_N(-\pi^2)$ and $K'_N(-\pi^2)$ are infinite if one does not use regularization, and are arbitrary otherwise.

Ab) One pair; infinite order normalization.

One can generalize the previous result. Let us define by $B_N^{2,q}$ the closed loops of nucleons having two vertices from which emerge external pairs, and q vertices between which internal pairs are exchanged; for example

$B^{2,2}$ $B^{2,3}$

The sum of all the « chains », constitued with any number of such closed loops $B_N^{2,q}$, defines a « General interaction kernel » $(\widetilde{NN})$, which is a function of the « General closed loop » $\widetilde{B}_N^2 = \sum_{q\to 0}^{\infty} B_N^{2,q}$ the value of which is $\widetilde{K}_N(p^2)$,

$$(\widetilde{NN}) = \frac{\gamma'}{1 - \gamma' \widetilde{K}_N} . \tag{N1.5}$$

Let us define also the « General normalized interaction kernel of the Yukawa theory » which is $iG^2\Delta'_F$; $\{\Delta'_F$ is the Dyson's function $(\Delta'_F)^{-1} = X(p^2)\cdot \cdot(1 - iG^2\widetilde{K}_N^R/X(p^2))\}$. In the same way as before, one generalizes the equivalence theorem (E), by determining the mesons equations in such a way that

$$iG^2\Delta'_F = (\widetilde{NN}) . \tag{N1.6}$$

It results also that the bare meson equation is a function of the infinite part of the function $\widetilde{K}_N$ and is linear.

Ba) L pairs; first order normalization.

To demonstrate the equivalence theorem (E), one must previously calculate the first order normalized kernels of Yukawa theory (UV_Y^1). To do this, one chooses the arbitrary constants of the K_u functions, in such a way that poles and residues of the (UV_Y^1) are the same as those of (UV_Y^0). But one disposes generally of $2L$ arbitrary constants to satisfy $L(L-1)$ conditions (L poles or meson masses, L^2 residues or coupling constants); therefore, except if there exist $L^2 + L - 2L = L(L-1)$ relations between the mesons constants one cannot normalize this theory to the first order. The R-theory gives these relations, one of which is the relation between the coupling constants $\sum_l G_u^l G_v^l = 0$ if $u \neq v$. However, even if these relations are fullfilled in lower approximation, they are not in higher approximations.

Bb) L pairs; infinite order normalization.

Let us define as $I_{uv}^{2,q}$ the irreducible closed loops which have the following

properties:

— They have 2 vertices from which emerge external pairs; the pairs carried by enclosing line are of u-type at one vertex and of v-type at the other.

— They have q other vertices between which internal pairs are exchanged.

— The irreducibility is the following topological property: Let x_u and x_v be the positions of the two « external » vertices in a two dimensional mapping; then, every open line crossing the loop $B_{uv}^{2,q}$ between these two points, and cutting lines of the loop only at the vertex points other than x_u and x_v, does not ever cross only one vertex.

Let us then define the functions $\widetilde{K}_{uv}$ represented by the sum of all these closed loops, $\widetilde{B}_{uv}^{2} = \sum_{q}^{\infty} B_{uv}^{2,q}$; if $u = v$ the smallest q is null and if $u \neq v$ the smallest q is two, or three when Furry's theorem applies.

The « General interaction kernels in R-theory », are then functions of the $\widetilde{K}_{uv}$; in a same way as in $(R\cdot 3c)$, one obtains

$$(\widetilde{UV}_F) = \gamma_{uv} + (\widetilde{UW}_F)\widetilde{K}_{wx}\gamma_{xv}\,, \tag{N1.7}$$

we denote here $\gamma_{uv} = ig_{uv}/2\pi$, g_{uv} being the coupling constant of the two pairs u and v (for an example of this calculation in the case of 2 pairs, see: V° Chap. VIII, F). Let us now define the « General normalized interaction kernel in the Yukawa theory », $(\widetilde{UV}_F)$ by the formula

$$(UV_F) = (UV_F^0) + (\widetilde{UW}_F)\widetilde{K}_{wx}^{R}(XV_F^0)\,, \tag{N1.8}$$

in which, (UV_F^0) denotes the « bare mesons interaction kernel »

$$(UV_F^0) = \sum_{l} iG_u^l G_v^l / X_l(p^2), \qquad X_l(p^2) = p^2 + \varkappa_l^2\,, \tag{N1.9}$$

and $\widetilde{K}_{wx}^{R}$ the general normalized closed loop functions, obtained from $\widetilde{K}_{wx}$. Let us explain this calculation. If the general closed loop functions, $\widetilde{K}_{uv}$, have the same types of divergences as the first order closed loops K_u used in previous sections, they contain $2\cdot L(L+1)/2$ arbitrary constants after a regularization of these expressions in the whole. One can thus determine, in normalization, all the constants, and this is equivalent to impose two conditions to every K_{uv}^{R} functions. As before, one has

$$\widetilde{K}_{uv} - \widetilde{K}_{uv}^{R} = F_{uv}\cdot(p^2 + b_{uv}) = \text{binomial } P_{uv} \tag{N1.10}$$

P_{uv} is the infinite part of $\widetilde{K}_{uv}$ before regularization; after it, F_{uv} and b_{uv} are arbitrary constants; P_{uv} is a pure imaginary quantity. The condition

$$(\widetilde{UV}_F) = (\widetilde{UV}_F)\,, \tag{N1.11}$$

determines the mesons of the equivalent Yukawa theory. After some manipulations of the equations (N1, 7, 8, 11) one gets

$$(UV_F^0) = \gamma_{uv} + (UW_F^0)P_{wx}\gamma_{xv} . \tag{N1.12}$$

It results that the field equations of mesons are linear functions of p^2. The bare meson field operator, Π^l is defined by

$$\sum_l G_u^l \Pi^l = \sum_v (UV_F^0) J_v , \tag{N1.13}$$

which gives the well known result when using the explicit form (N1.9).

As a conclusion, let us remark that, if the series of approximations of the unrenormalized closed loop expressions is strongly convergent, then P_{uu} $(u = v)$ are smaller than P_{uu} (the first order P_{uv} $(u \neq v)$ is of the same order of magnitude than second or third order P_{uu}). If one neglects them, one gets the relations between the coupling constants and the masses of the various mesons, underlined in Ba. This will perhaps provide an experimental test of the convergence of series expansions.

2. – On the Heisenberg theory of elementary particles [1].

It is useful to discuss some common features and differences between the Heisenberg theory and mine, since both are related to a realistic concept of elementary particles. The Heisenberg way starts from the only neutrino field, non-linearly coupled with itself through a Fermi type coupling, and calculates the various excited and compound states of this field (deductive way).

We use, on the contrary, an inductive approach:

— In a first step, mostly developped in (V^0) and (R), we determine the Fermi couplings between the various fermions, supposed elementary, in such a way that the resulting bosons are the experimentally known particles (photon, pions, ...).

— The second step will consist, when we suppose given the bosons and the « most elementary » fermions, in constructing the fermions mesons, in such a way that they are coupled by the Fermi coupling determined in step 1.

[1] I am now referring to a series of papers which were unfortunately unknown to me when I sent paper (R) to be published: W. Heisenberg: I - *Nachr. d. Göttinger Akad. Wiss.*, S. III (1953); II - *Zeits. f. Naturf.*, **9***a*, 292 (1954); III - To be published (in collaboration with F. Kortel and H. Mitter). I am very indebted to Prof. Heisenberg for giving me a preprint of his last paper (III), of which I discuss mostly here.

The solution of this problem, must indicate what are (or is) the most elementary fermion fields, and their coupling, (ν, e, p, n, or ν alone?).

Other differences appear for the treatment of the equations.

The Heisenberg equation and the associated method leads to finite results. Every closed loop is finite, because it must always be calculated with the use of a propagator $S(p)$ very strongly regularized as a consequence of the non-linear equation. The masses are calculated by using non-adiabatic methods, but, as pointed out in R. *bis*, Heisenberg is led to follow a way similar to that we used in (V°) and (R), to calculate the coupling constant (fine structure constant). A phenomenon analogous to our « Deutrino », (see V°, Chap. XII, XIII), appears also in his solution of the null mass boson problem (first term of equations 67 and 68 of III).

Now for a synthesis of our two theories (2). Let us postulate that, in the end, my way of approach will show that one can actually construct every particle with the only neutrino field, conveniently coupled with some Fermi type couplings (according to the Heisenberg idea). In this model, one must be able to show (step 2), that, starting from neutrino field and various boson fields given « a priori », one obtains all the fermions which are coupled according to the Fermi coupling calculated in step 1. But, this is equivalent to start from the general propagator of neutrino field, (more precisely from the fermion interaction kernel between two events where two pairs interact), which contains the various excited states of the neutrino.

The Fermi couplings that we have calculated in step 1, and that the step 2 must explain, have the following properties:

In all the cases when it was possible to compare two Fermi constants, they were found to have the *same absolute value*. This property has been indicated in (F) as resulting from the charged Fermi interactions measurements; but this is not the most convincing test, since the experiments give the value of the interactions and not the value of the coupling constants. The previous property results mostly

a) from the equality of magnitude of electric charge of p and e;

b) from the equality of magnitude of mesonic charge of p and n (3).

In two cases these constants have *opposite signs*.

(2) We give here some details about the arguments given in the paper (F).

(3) Let us remark here that the charge independence of nuclear forces, $(G^0_{nn})^2 = (G^0_{pp})^2 = -(G^0_{nn})(G^0_{pp}) = \frac{1}{2}(G^{\pm}_{np})^2$, that we have deduced in *Compt. Rend. Ac. Sc.*, t. 239, p. 1267 (1954) from the coupling

$$\sum_{i=\text{Lept}} g\Big\{(\bar{\psi}_n\gamma^5\psi_p)(\bar{\psi}_\nu\gamma^5\psi_i) + \big((\bar{\psi}_n\gamma^5\psi_n) - (\bar{\psi}_p\gamma^5\psi_p)\big)\cdot(\psi_i\gamma^5\psi_i \pm \bar{\psi}_\nu\gamma^5\psi_\nu)\Big\} + \text{c.c.}$$

in using the approximation of (R.5*a*1), results only from the fact that the Fermi constants of the pairs (nn) and (pp) have the same absolute value, and opposite signs. The com-

It results from the last property that step 2 leads to the existence of imaginary Fermi couplings (Ghost couplings), which appear necessarily associated with the convenient couplings. Then, either one must exclude the Heisenberg idea, or there must exists in nature some kind of « Ghost coupling exclusion principle ». I admited in (F) the second eventuality. The demonstration of this principle has to be formulated in accordance with the general formalism of quantum mechanics. Some consequences of this principle are investigated in (F).

3. – A new frame for mesons theory.

We shall briefly review the general restrictions imposed to the relations between the various mesons, as it results from the papers (R) and (F). We use a phenomenological classification of « absolute » (A) and « non absolute » (N.A.) selection rules (S.R.).

c) The Lorentz group S.R. - (A.).

b) The Isotopic spin S.R. - (N.A.).

A deeper reason for occurrence of isotopic spin quantum number, has been explained by the Fermi coupling ([3]): further calculations can now show the limits of its validity.

c) The allowed meson couplings - (A).

If one admits the postulate that there exist only the five types of Fermi couplings (S, V, T, A, PS) between certain fermions, then the resulting mesons can have only a few types of couplings with the fermion pairs; for example the PV coupling of PS mesons is excluded.

d) The Causality S.R. - (N.A).

The cut-off law, though quantitatively arbitrary, is however restricted by the causality condition, (R.5*b*). This restriction is more stringent when the particles resulting from decay have null masses; however, this S.R. does not apply when the interactions are non local. An interesting example is given by the μ-meson decay ([4]); one cannot have the process $\mu \to e + B^0 \to e + (\nu^* + \nu)$ (B^0 means some neutral boson) and, in the same order of the coupling constants, the process $\mu \to \nu + B^c \to \nu + (e + \nu^*)$ is allowed (B^c are some charged mesons). In high order processes, one can however obtain interactions between the neutrino pair and the $(e\mu)$ pair, by exchanging two charged bosons, which give non-

parison of the Fermi constants of neutral and charged couplings can be done by the calculation of the masses of charged and neutral pions, when the number of charged and neutral lepton pairs are better known.

([4]) We are indebted to Dr. L. Michel for an interesting discussion on this point.

local interactions. When the masses of the resulting particles are not null, the causality S.R. is only qualitative. We think that one must impute the preference decay $\pi \to \mu + \nu$ rather than $\pi \to e + \nu$ to a law of this type. Everything goes as if there existed some kind of « stepping law », which favours decays giving rise to the smallest lost of mass. Perhaps the Heisenberg theory can explain this fact quantitatively.

e) The Ghost S.R. - (N.A).

We indicated in (F) a tentative application of this law in a simplified model, which shows also an illustration of a complicated interaction which weakens the S.R. It is also a problem to determine if these S.R. can be expressed as a consequence of some phenomenological new quantum number (Gell-Mann's Strangeness, for example).

f) The Charge conservation - (A).

We introduce the charge conservation as a separate principle. But a realistic theory should be able to demonstrate this principle, in time.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

Cinetica e meccanismo delle reazioni di sostituzione nucleofila (*).

M. SIMONETTA

Istituto di Chimica Industriale dell'Università - Milano

(ricevuto il 19 Gennaio 1955)

Le reazioni di sostituzione semplice (nelle quali si ha lo scambio di un solo legame) si possono rappresentare in forma generale nel modo seguente:

$$RX + Y = RY + X .$$

In questa reazione si ha la rottura di un legame esistente e la formazione di un legame nuovo. Una prima distinzione si può fare in base al tipo di rottura del legame; possiamo avere scissione simmetrica o omolitica: R. . X oppure scissione asimmetrica o eterolitica: R|:X.

Il primo tipo è comune nelle reazioni in fase gassosa, mentre il secondo è caratteristico di molte reazioni di sostituzione in soluzione. Le reazioni di sostituzione eterolitica si dividono ancora in due classi a seconda che gli elettroni del legame che viene scisso rimangono a R oppure al gruppo sostituito X. Ciò dipende da due fattori: dal fatto che il gruppo che si stacca può avere la tendenza a separarsi con o senza gli elettroni del legame; e dal fatto che il gruppo sostituente può usare i suoi elettroni per attaccarsi al nucleo di R o usare il suo nucleo per attaccarsi agli elettroni di R. Le due classi di reazioni si chiamano rispettivamente nucleofile ed elettrofile. Possiamo allora costruire per le reazioni di sostituzione il quadro seguente.

a) Reazioni omolitiche $\quad Y. + R.|.X \rightarrow Y:R + X.$

b) Reazioni eterolitiche

1) nucleofile (S_N) $\quad Y:^- + R|:X \rightarrow Y:R + X:^-$

2) elettrofile (S_E) $\quad Y + R:^+|X \rightarrow Y:R + X^+$

(*) Conferenza tenuta a Napoli il 17 Ottobre 1954, in occasione della XLV Riunione della Società Italiana per il Progresso delle Scienze.

Siamo qui interessati alle reazioni di sostituzione nucleofila, indicate con il simbolo S_N. Queste reazioni sono state estensivamente studiate da C.K. Ingold, E. D. Hughes e dalla loro scuola: l'attenzione di questi autori si è rivolta in modo particolare al caso in cui la sostituzione avviene a un atomo di carbonio saturo, come succede, ad esempio, quando R indica un radicale alchilico. È stata provata l'esistenza di due meccanismi possibili:

un meccanismo a un solo stadio, o bimolecolare, S_{N_2}, ad esempio:

$$OH^- + R{-}Cl \rightarrow \overset{\sim 1/2-}{HO}....R....\overset{\sim 1/2-}{Cl} \rightarrow R{-}OH + Cl^-$$

e un meccanismo a due stadi, o monomolecolare, S_{N_1}, ad esempio:

$$R{-}Cl \rightarrow \overset{\delta+}{R}....\overset{\delta-}{Cl} \rightarrow R^+ + Cl^- \qquad \text{(lenta)},$$

$$R^+ + OH^- \rightarrow R{-}OH \qquad \text{(veloce)}.$$

L'introduzione in R di gruppi atti a cedere elettroni al centro di reazione favorisce il meccanismo S_{N_1} in confronto a quello S_{N_2}; il contrario accade con i gruppi che attirano elettroni.

Questo fatto è stato messo in evidenza molto chiaramente dallo studio della reazione di idrolisi dei cloruri alchilici in soluzione alcalina, come appare dalla tabella seguente, dove per i vari gruppi R sono indicati il tipo di meccanismo, l'ordine della reazione e la variazione della velocità passando da un gruppo all'altro:

meccanismo	S_{N_2}			S_{N_1}
gruppo (R)	CH_3-	CH_3-CH_2-	$(CH_3)_2CH-$	$(CH_3)_3C-$
ordine della reazione	2			1
velocità	— diminuzione →		— aumento →	

Si vede che la velocità di reazione passa per un minimo, nella regione di transizione tra i due meccanismi. Anche la natura del solvente, in modo speciale il suo potere ionizzante, può esercitare un'azione sensibile su queste reazioni in quanto un elevato potere ionizzante facilita un aumento di grandezza delle cariche elettriche nel passaggio dallo stato iniziale al complesso attivato, ne ostacola una diminuzione e ritarda la distribuzione di una data carica.

Tali azioni saranno più energiche nel caso di formazione o scomparsa di cariche elettriche che non nel caso della semplice distribuzione di cariche esistenti. Così negli esempi prima riportati per i due meccanismi S_{N_1} e S_{N_2} un aumento del potere ionizzante del solvente ritarderà blandamente il mecca-

nismo a uno stadio mentre avrà un'energica azione accelerante nel meccanismo a due stadi. Infine influirà sul meccanismo la natura dell'agente sostituente.

Se esaminiamo una serie di sostituzioni con diversi agenti nucleofili sullo stesso composto e nelle stesse condizioni, vediamo che, partendo con un reagente a grande affinità per il nucleo di carbonio con il quale la reazione procede secondo il meccanismo S_{N_2}, e considerando agenti di affinità via via minore, la velocità di sostituzione progressivamente decresce, sino a che si arriva al meccanismo S_{N_1}; dopo questo punto la velocità rimane naturalmente costante.

Volendo considerare i metodi sperimentali a disposizione per determinare il meccanismo effettivamente seguito nei vari casi occorre distinguere due tipi di sostituzioni nucleofile: solvolitiche e non solvolitiche. Nel caso delle reazioni non solvolitiche il gruppo sostituente viene da un reagente presente in concentrazione piccola e determinabile. Allora in generale a un meccanismo S_{N_2} corrisponde una cinetica del secondo ordine e a un meccanismo S_{N_1} una cinetica del primo ordine; quindi l'ordine della reazione osservato determina anche il meccanismo.

Nelle reazioni solvolitiche il gruppo sostituente viene da una notevole frazione del solvente, che è presente in concentrazione sostanzialmente costante durante tutto il corso della reazione. In prima approssimazione la reazione risulterà del primo ordine, indipendentemente dal meccanismo. Si ricorre allora ad altre tecniche, quali lo studio dell'effetto sulla velocità di reazione di cambiamenti strutturali nel composto o nel sostituente o di cambiamenti di solvente, o di aggiunta di sali; oppure lo studio approfondito dell'andamento cinetico della sostituzione o il suo corso stereochimico. Quest'ultimo metodo si è rivelato di grande importanza in quanto ha portato alla soluzione rigorosa del problema dell'inversione di Walden. Precisamente si sono potute formulare le seguenti due regole:

1) la sostituzione con meccanismo S_{N_2} comporta inversione di configurazione, indipendentemente da ogni particolare strutturale;

2) il meccanismo S_{N_1}, che passa attraverso uno ione carbonio, comporta racemizzazione, in generale accompagnata da un eccesso di inversione, a meno che non sia presente un gruppo che protegga la configurazione, nel qual caso la configurazione stessa viene in predominanza conservata.

Le reazioni di tipo S_{N_1} sono state definite anche « polimolecolari » in quanto è stato messo in evidenza in alcune reazioni di idrolisi la partecipazione di diverse molecole di acqua nel complesso attivato del primo stadio della reazione, quello di formazione dello ione carbonio; queste molecole di acqua avrebbero il compito di polarizzare il legame carbonio-cloro in modo da facilitarne la rottura. Recentemente è stata però riscontrata analoga partecipazione di un certo numero di molecole di solvente alla formazione del complesso attivato

anche per una reazione a meccanismo S_{N_2}, quale l'idrolisi dell'1-cloro-2-nitroetano in miscele acqua-diossano, acqua-acetone e acqua-acido formico.

La teoria delle reazioni di sostituzione nucleofila alifatica — che è stata qui riassunta nelle sue linee più generali — è rimasta fino ad oggi a uno stadio esclusivamente qualitativo. Ciò è dovuto in parte alla scarsa applicazione che hanno avuto finora i metodi della meccanica quantistica ai composti alifatici.

Passando ad esaminare le reazioni di sostituzione nucleofila aromatica, mentre le idee fondamentali della teoria per i composti alifatici sono ancora valide, vedremo che si può giungere a una trattazione semiquantitativa del problema, applicando, con opportune semplificazioni, il metodo degli orbitali molecolari. Salvo poche eccezioni, delle quali il più cospicuo esempio sono le decomposizioni di alcuni ioni diazonio, le sostituzioni nucleofile aromatiche avvengono secondo il meccanismo a uno stadio, S_{N_2}, e quindi limiteremo le nostre considerazioni a questo caso.

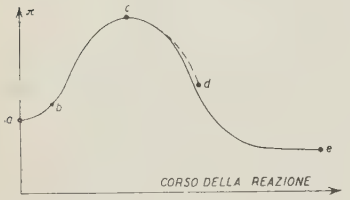


Fig. 1. – Curva della variazione d'energia degli elettroni π durante il corso della reazione.

Le variazioni di energia libera durante il corso della reazione si possono rappresentare con una curva come in fig. 1.

In questa curva il punto a corrisponde allo stato iniziale, cioè ai reagenti separati, c ai prodotti della reazione separati e c al complesso attivato. L'energia libera di attivazione, misurata dal dislivello fra c e a si può scrivere nella forma

$$\Delta F = \Delta\pi + \Delta\Sigma - T\Delta S\,,$$

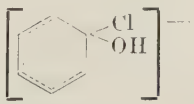
dove $\Delta\pi$ indica la differenza nelle energie degli elettroni π, $\Delta\Sigma$ la differenza nelle energie degli elettroni σ, nell'energia termica e di punto zero (dovute al moto dei nuclei) e nelle energie di solvatazione, e ΔS è l'entropia di attivazione. Se ora riduciamo il nostro problema al confronto tra la reattività di due composti aromatici 1 e 2 di struttura simile o tra due posizioni 1 e 2 dello stesso composto, le differenze degli altri termini scompaiono rispetto alle differenze $\Delta\pi$ e pertanto anzichè confrontare ΔF_1 e ΔF_2 possiamo paragonare $\Delta\pi_1$ e $\Delta\pi_2$: vale a dire possiamo sostituire nella figura all'ordinata F l'ordinata π.

Per quanto riguarda lo stadio a sappiamo che il metodo degli orbitali molecolari (M.O.), nella sua approssimazione data da una combinazione lineare di orbitali atomici (L.C.A.O.), permette di calcolare almeno approssimativamente l'energia degli elettroni π in una molecola aromatica e la concentrazione elettronica ai vari nuclei, ma per lo stadio c la situazione è diversa e occorre introdurre qualche altra semplificazione. I metodi semplificati attualmente in uso sono due: il metodo della molecola isolata, che sostituisce allo stadio c

lo stadio *b* (corrispondente al primo avvicinarsi del reagente alla molecola aromatica) e il metodo della localizzazione, che sostituisce allo stadio *c* lo stadio *d* (cioè al complesso attivato un modello ideale accessibile ai nostri calcoli).

Nella approssimazione della molecola isolata, la reattività di una posizione in una reazione con un ione si considera essere determinata dalla concentrazione di elettroni π presenti in quella posizione al momento della reazione. Se il reagente è uno ione negativo (come accade di solito nel caso di sostituzioni nucleofile in nuclei aromatici) gli elettroni π del sistema coniugato tenderanno ad allontanarsi dallo ione quando questo si avvicina. Teoricamente si può ottenere lo stesso effetto con un aumento dell'integrale di Coulomb dell'atomo nella posizione attaccata: si può allora assumere che l'effetto prodotto dall'avvicinarsi dello ione sia un aumento dell'integrale di Coulomb dell'atomo al quale lo ione si avvicina. Dalla teoria generale del metodo degli orbitali molecolari è noto che $\delta E \approx q_r \delta\alpha_r$, dove q_r è la densità di elettroni π all'atomo r e α_r è l'integrale di Coulomb per lo stesso atomo. Poichè $\delta\alpha_r$ è positivo, il più piccolo aumento di E si avrà per il più piccolo valore di q_r, cioè più piccola è la densità di carica q_r nel punto sede della sostituzione, maggiore è la reattività nucleofila iniziale in quella posizione.

Nella approssimazione della teoria della localizzazione, come modello del complesso attivato si prende uno ione in cui il reagente è legato con un legame covalente σ all'atomo di carbonio attivato, mentre il gruppo o l'atomo che vanno sostituiti sono tuttora legati allo stesso atomo di carbonio mediante legame covalente σ (in realtà si dovrebbero assumere dei legami parziali). Ad esempio, nel caso del clorobenzolo attaccato da uno ione ossidrile, il modello del complesso attivato sarà il seguente:



L'atomo di carbonio sede della sostituzione, originariamente in uno stato di ibridazione sp^2, è assunto essere in uno stato prossimo a quello di ibridazione sp^3, e non fa più parte del sistema coniugato. La differenza in reattività tra molecole simili o tra le diverse posizioni di una molecola è data — in questa approssimazione — dalla differenza nella variazione di energia del sistema di elettroni π nel modello assunto per il complesso attivato (posizione *d* nella curva di fig. 1) e nello stato iniziale (posizione *a*).

Entrambe le teorie sono state ripetutamente saggiate su vari derivati benzenici e sempre con esito positivo. Recentemente il metodo è stato esteso anche a derivati della naftalina. Precisamente è stata determinata la velocità di reazione dell'1-cloro-2-nitronaftalina (*A*) dell'1-cloro-4-nitronaftalina (*B*) e della 1-nitro-2-cloronaftalina (*C*) con idrato potassico o con metilato o etilato

sodico in solventi vari e a diverse temperature. In ogni caso l'ordine di reattività è risultato il seguente: $A > B > C$.

Il metodo della molecola isolata ha portato ai seguenti risultati: nell'1-nitronaftalina $q = 0.792$ (in unità elettroniche) in posizione 2 e $q = 0.632$ in posizione 4; nella 2-nitronaftalina $q = 0.566$ in posizione 1 (nell'esecuzione del conto sono stati adottati, per i parametri necessari, i valori usati da WHELAND e PAULING per il nitrobenzolo ed è stata trascurata la coniugazione con gli elettroni p del cloro e l'effetto induttivo di questo atomo). Similmente il metodo della localizzazione ha dato i risultati che seguono: per A, $\Delta W = -0.997$ (in unità $\beta = -20$ kcal/mole); per B, $\Delta W = -1.1682$; per C, $\Delta W = -1.2468$. Come si vede chiaramente dal grafico di fig. 2, i risultati sperimentali e quelli teorici sono in ottimo accordo.

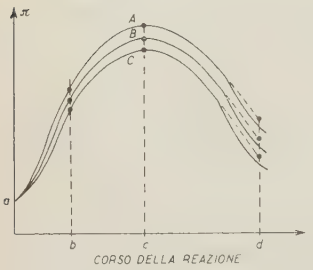
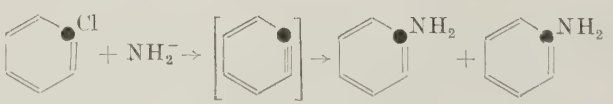


Fig. 2. – Aspetto qualitativo delle curve di energia potenziale: A = 1-nitro, 2-cloronaftalina; B = 1-nitro 4-cloronaftalina; C = 2-nitro 1-cloronaftalina.

Alcuni mesi fa si è scoperto che le reazioni di sostituzione aromatiche mediante un agente nucleofilo possono avvenire anche con un diverso meccanismo, che possiamo definire di eliminazione-addizione. J. D. ROBERTS, nel 1953, ha pubblicato una breve nota sui processi di amminazione basica degli alogenuri arilici, notando che queste reazioni sono spesso accompagnate da processi di «riarrangiamento»; ad esempio da orto o metacloroanisolo si ottiene la meta-anisidina, mentre dal composto para si ottiene una miscela di para e meta derivato. Inoltre queste reazioni non hanno luogo se in posizione orto all'alogeno da sostituire non c'è almeno un atomo di idrogeno. Questi fatti suggerivano un processo di deidroalogenazione da parte degli ioni NH_2^-. ROBERTS ha condotto la stessa reazione sul clorobenzolo contenente un atomo ^{14}C (quello al quale è attaccato il cloro). Operando con potassioammide in ammoniaca liquida a -33 °C il cloro viene sostituito con grande rapidità e e l'anilina così ottenuta possiede il gruppo amminico per il 50% attaccato all'atomo di carbonio marcato, e per il 50% in posizione orto a tale atomo.

C. K. INGOLD, anche in base a considerazioni sugli stati eccitati delle molecole di benzolo e acetilene, ha formulato per la reazione il seguente schema:



con la formazione intermedia di un «benzino».

Dopo aver discusso le reazioni di sostituzione nucleofila alifatica e aroma-

tica, consideriamo ora un caso in certo senso intermedio: quello delle reazioni di sostituzione nucleofila nelle catene laterali di derivati benzenici. Per le reazioni di composti di questo tipo è stata formulata su basi puramente empiriche da HAMMETT una teoria, della quale recentemente è stata proposta da JAFFÈ una interpretazione su basi quantomeccaniche. Questa teoria riguarda le reazioni in genere nelle catene laterali dei derivati benzenici, e qui ne sarà considerata la sua applicazione al caso particolare di reazioni di sostituzione nucleofila. Secondo HAMMETT nel caso di derivati benzenici para e meta sostituiti, se si paragonano due serie di costanti di equilibrio o di velocità pertinenti a reazioni che differiscano o per la natura del gruppo reagente o per il cambiamento che esso subisce per effetto della reazione, si trova che esiste una semplice relazione quantitativa tra queste costanti. La stessa cosa si ha quando si paragonino fra loro i membri di una stessa serie di reazioni. (I membri di una serie sono identici in quanto alla reazione, ma differiscono per i gruppi sostituiti in meta o in para al gruppo reagente).

Una serie può riferirsi, ad esempio, alle costanti di ionizzazione degli acidi benzoici meta o para-sostituiti: un'altra alle costanti di velocità di idrolisi di esteri degli stessi acidi. Se si riportano in un diagramma logaritmico le costanti di una serie in funzione delle costanti dell'altra, si osserva che i punti rappresentativi si dispongono con grande approssimazione su di una retta. Punti rappresentativi di derivati alifatici o benzenici orto-sostituiti si scostano invece decisamente dalla retta precedente e si dispongono a caso sul diagramma. Per la retta si può scrivere l'equazione:

$$\log K_h = \varrho \log K_i + A \,, \tag{1}$$

in cui K_h è la costante di velocità di idrolisi degli esteri, K_i la costante di equilibrio della dissociazione acida, e ϱ e A sono costanti.

Analoga relazione lineare vale praticamente per tutte le reazioni delle catene laterali dei derivati benzenici meta e para sostituiti e in particolare per le reazioni di sostituzione nucleofila.

L'equazione (1) si può trasformare nel modo seguente:

$$\log K = \varrho(\log K_i - \log K_i^0) + (A + \varrho \log K_i^0) \,,$$

dove K è una generica costante di velocità o di equilibrio e K_i^0 è la costante di ionizzazione dell'acido benzoico. Necessariamente $A + \varrho \log K_i^0$ è uguale a $\log K^0$, dove K^0 è la corrispondente costante di velocità o di equilibrio per la molecola non sostituita. Definendo allora $\sigma = \log K_i - \log K_i^0$ si ha:

$$\log K - \log K^0 = \sigma\varrho \,. \tag{2}$$

La costante σ è per definizione fissata dal sostituente ed è indipendente dalla reazione cui si riferisce la costante K: la costante della reazione ϱ è la medesima per tutti i sostituenti e dipende solo dalla reazione in studio. Per tutt

i sostituenti per i quali si conosce con sufficiente esattezza la costante di ionizzazione del corrispondente acido benzoico sostituito, si ha immediatamente il valore di σ dalla definizione stessa. Per qualunque reazione per la quale si abbiano dati per un sufficiente numero di derivati meta e para sostituiti si calcola ϱ riportando sul diagramma il $\log K$ in funzione di σ; con i valori di ϱ così ottenuti si possono calcolare altri valori di σ non ottenibili direttamente. HAMMETT ha riportato tabelle di valori di σ per i principali sostituenti in meta e in para e dei valori di ϱ per oltre cinquanta serie di reazioni.

Poichè l'effetto di un sostituente è essenzialmente quello di produrre diversi cambiamenti di carica elettronica sui diversi atomi di carbonio dell'anello benzenico, è logico cercare un collegamento tra i valori delle costanti σ di HAMMETT e le densità elettroniche nei vari punti della molecola. Il calcolo delle densità elettroniche in nuclei benzenici sostituiti è stato eseguito per via teorica da WHELAND e PAULING applicando il metodo degli orbitali molecolari nella sua approssimazione standard di combinazione lineare di orbitali atomici. Le difficoltà inerenti all'applicazione di tale metodo di calcolo a composti contenenti eteroatomi sono essenzialmente quelle di assegnare gli opportuni valori agli integrali coulombiani degli eteroatomi presenti nella molecola, agli integrali di risonanza carbonio-eteroatomo e di tener conto dell'influenza dell'eteroatomo sugli atomi di carbonio vicini. Queste difficoltà sono state superate da WHELAND e PAULING e dagli autori che successivamente si sono occupati di questo problema assegnando valori empirici alle grandezze in questione.

Successivamente è stato messo in evidenza che la maggior parte dei calcoli eseguiti non era sensibile alla scelta dei valori assegnati a detti parametri. Una grandezza estremamente sensibile a tali valori è stata recentemente identificata da JAFFÈ nei valori degli incrementi delle densità elettroniche nelle posizioni para e meta al sostituente e soprattutto del loro rapporto.

Consideriamo un composto avente un sostituente X in posizione meta o para rispetto a una catena laterale Y. L'effetto del sostituente X in meta o in para a Y in una reazione della catena laterale Y è presumibilmente il risultato dell'influenza di X sulla densità elettronica del centro reattivo. Il sostituente modifica la distribuzione degli elettroni π nell'anello benzenico; il cambiamento di carica dell'atomo di carbonio al quale è attaccata la catena è trasmesso al centro reattivo in modo che dipende dalla natura della catena laterale Y. La validità generale dell'equazione (1) dimostra che il cambiamento prodotto da X nell'atomo di carbonio al quale è attaccata la catena è essenzialmente indipendente dalla natura di Y. (Naturalmente si escludono i casi in cui Y fa parte del sistema di risonanza). Perciò si può assumere che un analogo cambiamento di densità elettronica sarà prodotto da X nella molecola corrispondente nella quale ad Y sia sostituito un atomo di idrogeno. Segue che gli incrementi di densità elettronica nelle posizioni meta e para dei benzoli monosostituiti dovrebbero essere una funzione monotona dei corri-

spondenti valori di σ; e in buona approssimazione tale funzione si può ricondurre a una proporzione diretta tra i valori di σ e gli incrementi di densità elettronica.

JAFFÈ ha eseguito i calcoli necessari per determinare i valori dei parametri nel caso di diversi derivati benzenici monosostituiti del tipo più semplice, cioè quelli che contribuiscono con un solo orbitale π al sistema di risonanza. Nel caso delle posizioni orto la densità elettronica è calcolata correttamente con gli stessi parametri; i valori delle costanti di velocità sono però influenzati da effetti sterici dei quali occorre tener conto almeno qualitativamente se si vogliono includere questi derivati nella teoria. Il successo dei lavori di JAFFÈ dimostra la possibilità di collegare le densità elettroniche nei benzeni monosostituiti coi valori σ di HAMMETT almeno nei casi in cui la variazione di tali densità prodotta dal sostituente sia piccola, ed è buona garanzia che l'ipotesi fondamentale di una dipendenza monotona delle costanti σ dalle densità elettroniche è essenzialmente corretta. Precisamente JAFFÈ ha stabilito la relazione $\sigma = -15.5\,\Delta q$ per cui, nel caso di una reazione per la quale sia stabilito sperimentalmente, mediante altri derivati, il valore di ϱ, è possibile, dal valore di Δq calcolato col metodo M.O. per un nuovo sostituente, predire il valore della costante di velocità.

Una nuova conferma della relazione tra i valori σ di HAMMETT e la concentrazione di elettroni all'atomo di carbonio al quale è saldata la catena è stata recentemente ottenuta da GUTOWSKI e collaboratori misurando lo schermo del momento magnetico del nucleo ^{19}F nel fluorobenzene e in alcuni fluorobenzeni sostituiti, e stabilendo l'esistenza di una relazione tra le loro costanti di schermo δ e le costanti σ. La relazione tra i valori di σ e la concentrazione elettronica è così confermata da risultati di misure cinetiche, di misure magnetiche e di calcoli quantomeccanici.

Come esempio di applicazione della teoria di Hammet-Jaffè a reazioni di sostituzione nucleofila, esaminiamo il caso della idrolisi del cloruro di benzile e dei suoi derivati orto, meta e para metossi-sostituiti in miscele acqua-acetone. Le densità elettroniche nelle varie posizioni dell'anisolo sono:

$$\text{orto: } 1.036\,, \qquad \text{meta: } 0.992\,, \qquad \text{para: } 1.017\,.$$

Inoltre è

$$\sigma_m = 0.115\,, \qquad \sigma_p = -0.268\,.$$

Le reazioni solvolitiche, tutte rigorosamente del primo ordine, decorrono secondo il meccanismo S_{N_1} e la reattività dei vari composti decresce nell'ordine:

$$\text{para} > \text{orto} > (\text{non sostituito}) > \text{meta}\,,$$

cioè nello stesso ordine in cui decrescono le concentrazioni elettroniche all'atomo di carbonio al quale è legata la catena laterale, eccetto che per l'inversione orto/para dovuta ad effetti sterici.

Passando al cloruro di β-feniletile e ai suoi derivati metossi-sostituiti, è stata determinata la costante k_2 di velocità della reazione con KOH in soluzione acqua-alcool. I valori di k_2 (in $h^{-1}/mole^{-1}\,l^{-1}$) sono riassunti nella seguente tabella:

	60 °C	70 °C		60 °C	70 °C
non sostituito	0.340	1.080	meta . . .	0.510	1.410
orto . . .	0.0725	0.224	para . . .	0.125	0.370 .

In questo caso le reazioni sono del secondo ordine e il meccanismo è S_{N_2} (in realtà è presente anche la reazione di eliminazione a meccanismo E_2 con formazione di oleofina); e i dati, in ottimo accordo con la teoria, permettono di valutare graficamente ϱ (fig. 3).

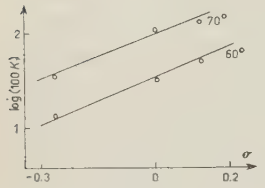
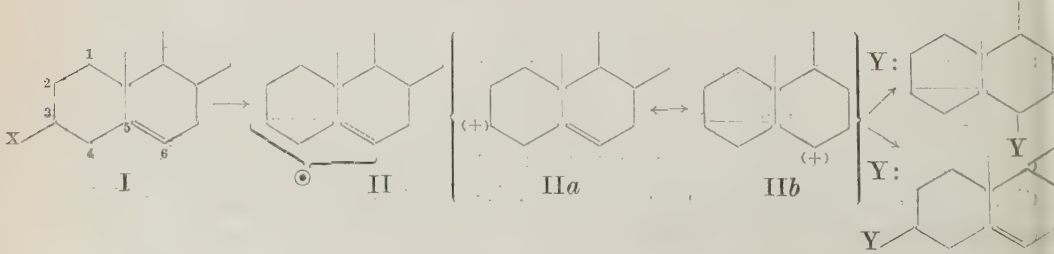


Fig. 3. – Curva log k_2, per la reazione fra i cloruri di metossifenile e idrato potassico in alcool acquoso: $\varrho_{70^0} = 0.0156$, $\varrho_{60^0} = 0.0164$.

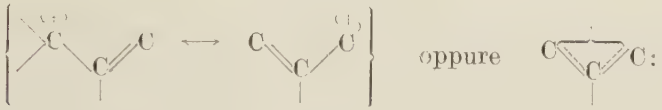
Se consideriamo la reazione di sostituzione in assenza di KOH troviamo a 95 °C i seguenti valori di k_1 (in h^{-1} perchè la reazione è del primo ordine): cloruro di feniletile 0.0034, cloruro di para-metossifeniletile 0.0105. Si vede come passando dal meccanismo S_{N_2} al meccanismo S_{N_1} l'ordine di reattività dei due composti è invertito.

Ci sono poi delle reazioni il cui andamento cinetico non rientra in nessuno degli schemi già citati, in quanto alcuni fattori che generalmente hanno importanza secondaria e quindi vengono trascurati nelle schematizzazioni adottate, assurgono ad un ruolo predominante determinando così l'andamento cinetico della reazione. Come primo esempio di queste reazioni particolari citeremo il caso delle interazioni 1,3 o della risonanza cosidetta omoallilica.

È stato notato che i derivati colesterilici hanno una notevole velocità di ionizzazione, attribuibile alla delocalizzazione della nuvola elettronica del gruppo olefinico adiacente al centro di reazione nello stadio che determina la velocità. Questa partecipazione del gruppo olefinico 5,6 al processo di sostituzione dà luogo ad uno ione II: che si può scrivere come risonante tra le strutture II*a* e II*b*, con parziale legame 3,5 e indebolimento del legame 5,6. Questa specie intermedia reagisce con le specie nucleofile a C_6 per dare 3,5 ciclosteroidi (III) o a C_3 per dare derivati colesterilici (IV).

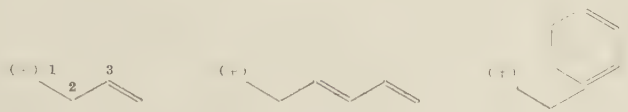


Il termine omoallilico è stato dedotto in base all'analogia col caso dello ione allilico:



rispetto allo ione allilico si ha l'interposizione di un gruppo $—CH_2—$ tra il centro cationico e il sistema di elettroni π.

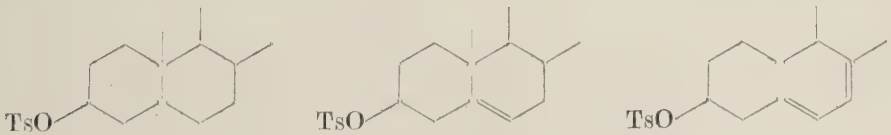
Per alcuni casi semplici, quali i sistemi V, VI VII



V VI VII

è stato possibile calcolare l'energia di stabilizzazione competente alla configurazione più stabile, nella quale si ha equilibrio tra la tendenza del centro cationico ad avvicinarsi al sistema olefinico con conseguente aumento dell'energia di risonanza, e la resistenza a diminuire l'angolo tra i legami o tra gli atomi di carbonio 1,2 e 2,3 con conseguente diminuzione dell'energia di tali legami.

I valori trovati per tale energia sono: 6.2; 9.6 e 4.2 kcal/mole per V, VI, VII rispettivamente. L'ordine di stabilizzazione dienile > vinile > fenile trovato per questi gruppi in posizione β è lo stesso che per l'ordinaria coniugazione di questi gruppi con un legame olefinico. Ma ciò che più interessa in questo momento è l'ordine degli effetti di aumento di velocità dei β sostituenti nelle reazioni di solvolisi. Per esempio, le velocità di solvolisi dei paratoluensolfonati di colestanile VIII, colesterile IX e deidrocolesterile X sono nel rapporto 1 : 100 : 3000. Tali valori corrispondono a una stabilizzazione degli stati



VIII IX X

di transizione per IX e X di circa 3 e 5 kcal/mole anche se l'orientamento spaziale non è l'ideale per la sovrapposizione degli orbitali π.

Un altro caso particolare che discuteremo è la reazione di sostituzione del cloro con iodio nei derivati feniletilici. Conant e Kirner hanno determinato le costanti di velocità del secondo ordine per le reazioni di sostituzione a mec-

canismo S_{N_2}:

$$R—Cl + KJ = R—J + KCl$$

per i composti $C_6H_5(CH_2)_nCl$ con n variabile da 1 a 7. I valori delle costanti cadono bruscamente passando dal composto con $n=1$ a quello con $n=2$: risalgono alquanto per $n=3$ per tornare a scendere per $n=4$ e rimanere quindi costanti (ed eguali al valore trovato per i derivati alifatici a catena normale) al crescere di n. Recentemente lo studio cinetico di questa reazione è stato esteso ai derivati orto, meta e para metossi-sostituiti del cloruro di benzile, di β-feniletile e di γ-fenilpropile. Si è trovato che un aumento della concentrazione elettronica al centro di reazione favorisce la reazione nel caso dei derivati benzilici e fenilpropilici, mentre la ritarda nel caso dei derivati feniletilici, e viceversa per una diminuzione di tale concentrazione.

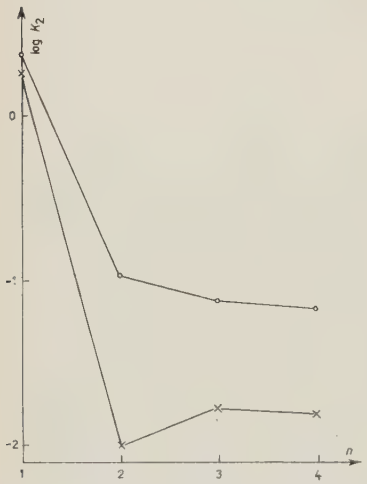


Fig. 4. – Logaritmi delle costanti di velocità sperimentali (×) e ideali (○) in funzione del numero di atomi di carbonio della catena laterale.

Come è noto, entrambi gli effetti sono prevedibili nelle reazioni a meccanismo S_{N_2}; il diverso tipo di comportamento nei singoli casi si può attribuire al fatto che il maggiore contributo al valore attuale dell'energia di attivazione può venire sia dal processo di parziale rottura del legame esistente che dal processo di parziale formazione del legame nuovo.

Queste considerazioni e le analogie strutturali esistenti tra questo caso e l'analogo esempio di diminuzione di reattività passando dal cloruro di butile terziario al cloruro di neopentile, inducono a pensare che la causa della forte diminuzione di velocità sia da ricercarsi in fenomeni sterici. Si può peraltro calcolare approssimativamente, secondo un metodo dovuto a Ingold, Hughes e Dostrovsky, l'incremento di energia prodotto da impedimento sterico nel passaggio da composto iniziale a complesso attivato nel caso dei derivati benzilici, feniletilici e fenilpropilici: i risultati sono rispettivamente 0.2 kcal/mole, 1.4 kcal/mole e 0.9 kcal/mole (quest'ultimo valore rimane costante per gli omologhi superiori della serie).

È ora possibile, togliendo alle energie di attivazione totale la parte « sterica », calcolare i valori delle costanti di velocità « ideali » della reazione, vale

a dire i valori che tali costanti avrebbero se fosse possibile abolire le forze di compressione tra gli atomi non legati che, nel complesso attivato, vengono a trovarsi troppo vicini fra loro: i risultati appaiono in fig. 4, dalla quale si vede che via via che ci si allontana dal gruppo fenilico la reattività del cloro decresce regolarmente per raggiungere il valore costante dei derivati alifatici.

Si raggiunge così una spiegazione conforme all'apparente anomalia presentata dalle costanti sperimentali.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

The Differential-Space Theory of Quantum Systems.

N. WIENER

Massachusetts Institute of Technology

A. SIEGEL (*)

Boston University

(ricevuto il 7 Febbraio 1955)

CONTENTS. — 1. The Question of « Hidden Variables » in Quantum Mechanics. – 2. Individual and Ensemble. – 3. The distinction Between Operators and Observables. – 4. Dynamical Specification and Dynamics in Differential-Space Theory. – 5. Summary of Properties of Differential Space; Restatement of the Basic Theorem of Differential-Space Quantum Theory. – 6. Statistical Part of the Theory.

1. – The Question of « Hidden Variables » in Quantum Mechanics.

Since the very earliest period, the physical interpretation of quantum mechanics has been a subject of controversy. As the discussion progressed, it gradually became centered around the conflict between two well-defined opposite points of view [1-5]. Although hitherto neither has obtained a clear-cut logical victory, most physicists have tended to resolve their doubts in favor of one

(*) Research of this author supported in part by U. S. Office of Naval Research.

(1) The history of the controversy is well summarized (with references) in essays by N. BOHR (pp. 201-241) and A. EINSTEIN (pp. 665-688) in *Albert Einstein: Philosopher-Scientist* (Evanston, Illinois (1949)).

(2) A. EINSTEIN, B. PODOLSKY and N. ROSEN: *Phys. Rev.*, **47**, 777 (1935).

(3) N. BOHR: *Phys. Rev.*, **48**, 696 (1935).

(4) W. FURRY: *Phys. Rev.*, **49**, 393 (1936).

(5) A. EINSTEIN: *Physics and Reality*, § 5, in *Journ. Frank. Inst.*, **221**, 313 (1936), reprinted in A. EINSTEIN: *Out of My Later Years* (New York, 1950).

of them. We shall give a résumé of this discussion, as the most convenient way of defining the terms needed for presenting our theory (6).

The more generally accepted view has been that identified mainly with NIELS BOHR. BOHR adapts the physical picture to the schematic experimental situation, excluding any physical concepts that do not correspond strictly to experimental elements, even if there is no actual disagreement involved. Thus, in a situation in which the position of a particle is sharply defined, the very concept of momentum is excluded, since the wave function displays a wide dispersion of momentum values.

The principal advocate of the opposing view has been ALBERT EINSTEIN. To use Einstein's characterization, quantum mechanics (which may be considered a formal expression of Bohr's point of view) is not a « complete theory »: elements of physical reality, such as momentum in the above example, may exist and have definite values, even though unsharply defined according to the wave function.

Since advocates of the first view maintain that the quantum-mechanical description of reality is complete, it would follow that each individual of a number of identically prepared maximally defined (subject to the indeterminination relations) systems is *completely* represented as to it sphysical state by the same wave function. The fact that a measurement of a variable of which this wave function is not an eigenfunction will yield a variety of values for different individual systems is not, according to BOHR, to be used as an argument for supposing that there exists a theoretically definable difference between these systems prior to the measurement. Indeed, such a conclusion is easily shown not to be a necessary one: the variable which was sharp in the wave function and the variable finally measured correspond to mutually exclusive possibilities of definition experimentally; the process of measuring the latter includes an unobservable and uncontrollable disturbance of the value of the former, making its value different (7) in each case. This accounts for the difference between systems after the measurement, without assuming a difference before measurement. But BOHR replaces this negative conclusion (lack of necessity for assuming a theoretically definable difference prior to measurement) by a positive one. The differences between different systems with regard to the second variable are, so to speak, created at the moment of its measurement and do not exist previously. Not that all had identical values, rather one must say that the second variable (within more or less broad limits depending on

(6) The theorem that is basic to the present work was given in an earlier paper by the authors, N. WIENER and A. SIEGEL: *Phys. Rev.*, **91**, 1551 (1953). We refer to this paper here as WS.

(7) Actually this variable becomes *undefined* after the measurement, as was the other before; see below.

the particular wave function) has no clear-cut prior existence; not only its values, but the second variable as a defined concept, are first created by the act of measuring it. The logic of the assertion that the systems are different after the measurement, even though identical before, is protected by the fact of the unobservable disturbance of the first variable; this is the essential rôle of the indeterminacy principle.

Without questioning the ability of these concepts to account for the phenomena, EINSTEIN has insisted that it has not been proved that physical theory is constrained from examining the area of indeterminacy which BOHR maintains to be inscrutable even in concept. He asserts, in effect, that when there exists a difference between measured values of a variable in two systems, there esists an in principle physically definable difference between them. It is not necessary to deny the undeniable disturbance sometimes accompanying the act of measurement in order to assert this; indeed, the difference can be demonstrated within the very context of quantum mechanics itself, even when there is no disturbance ([2]). EINSTEIN in effect challenges Bohr's transformation, described above, of a negative into a positive statement.

As we shall develop it in this paper, the differential-space theory carries the controversy sketched above to a new stage of concreteness. This is because it realizes the formal program of EINSTEIN: namely, a physical picture in which a collection of identically prepared (subject to the indeterminacy principle) systems is represented by an ensemble of different, precisely defined systems so that the dispersion in values of measured quantities has its counterpart in the variety of properties of the systems in the ensemble. On the other hand, the theory and current experimental information are in too provisional a state for these facts to constitute in any sense a refutation of Bohr's point of view. That is, it is entirely possible that the individual-description concept formally furnished by the differential-space theory may turn out to be physically vacuous.

2. – Individual and Ensemble.

It is a time-honored fact that assigning (in imagination) sharp values to two conjugate variables of a system, and doing nothing more, will result in conflict with the experimental facts. In a two-slit diffraction experiment with electrons ([3]), the positional undefinability-in-principle of the electron in the Bohr picture serves an indispensable logical purpose. If one could think of the electron as defined within a region of extent less than the separation of the slits (assuming it to be experimentally not so well defined) one would be

([3]) See L. I. SCHIFF: *Quantum Mechanics* (New York, 1949).

able to assert (in thought) that the electron had certainly gone through one of the slits, and would therefore be unable to account for the influence of the other slit on the experimental results (the results show a diffraction pattern of structure dependent on the separation of the slits, since the electrons are by hypothesis experimentally so prepared as to give such a pattern — in particular, their experimental range of position definition is broader than the separation of the slits).

Referring to the first sentence of the preceding paragraph, it is then evident that in order to realize a description of quantum systems in which all observables are defined one must do « something more. » The classic way to do this « something more » is to add further determining — « hidden » — variables to the system; these can then serve the all-important purpose of expressing the influence of the « other slit » even when the particle is hypothesized (without experimental verification) as not passing through it.

A theory of hidden variables designed to explain quantum mechanics must contain two basic new features. The first feature, logically speaking, is concerned with the hidden variables of a single completely described system (in EINSTEIN's, not BOHR's sense). This we shall refer to as its *dynamical* part. The other essential part of such a theory is its *statistical* part. This has the following function: to show how ensembles of completely described systems obeying the postulated dynamics can be so constructed as to have the same statistical properties as are expressed by a given quantum-mechanical wave function. (If nature is correctly described by such a theory, individual events as described by the conventional quantum mechanics are *crypto-deterministic* (9), i.e., deterministic in themselves, but with an imprecision in their definition corresponding to an imprecision in their experimental measurement.)

The program stated in the preceding paragraph implies certain necessary conditions on its realization, tied up with the intuitive meanings our minds attach to the terms « deterministic » and « statistical ». The following is what we feel to be a « fair » or « reasonable » formulation of these conditions.

1) The individual system is deterministic, i.e., its present state (the present values of all measurable variables) determines its future state (the values of all measurable variables at any future time). The system must of course be closed if this is to apply. Open systems must first be combined with all systems with which they interact (or be subjected to some equivalent procedure), making them closed, before they can be deterministic. This condition concerns only what we have called the *dynamical* part of the theory.

The remaining two conditions refer to the *statistical* part. It is assumed, following Gibbs, that all statistical predictions — i.e., predictions regarding the distribution of values of physical properties over a large number of identically.

(9) This term is due to E. T. WHITTAKER: *Proc. Roy. Soc.*, **55**, 459 (1943).

but not (in the dynamical sense) precisely, defined systems — are to be obtained from an ensemble of dynamically precisely defined systems, such that a measure (or probability, or weight) is given for any suitably defined and restricted set of values of the variables. This measure is assumed to be the probability, relative to the ensemble, of the subensemble of all systems belonging to the ensemble and having values of their variables in this set. Since the ensemble is a mathematical representation of the set of actual systems observed, and one cannot isolate a negative number of systems, one infers the second condition.

2) The measure of a suitably defined subensemble is never negative ([10]).

The correspondence between the ensemble and the set of actual observed systems is presumed to be preserved with the passage of time. That is, the systems of the ensemble duplicate the time-behavior of the actual observed systems. The third condition is the formal expression of this idea.

3) The measure of a subensemble corresponding to a given set of variables is conserved an this set changes according to the natural motion (dynamics) of the variables contained in it (conservation of probability, or preservation of measure, in time; in classical statistical mechanics, the proof of this theorem rests on Liouville's theorem).

For a physicist, it appears logical to explain such a theory as ours, containing a new dynamics and a new statistical formalism, by starting with the dynamics. This procedure we adopt here. However, the dynamics and the statistics are so interwoven in the complete theory that this division is somewhat artificial. We therefore think it proper to point out in advance that our discussion of the dynamics suffers from an appearance of aridity which is due (apart from any defects in presentation) to this separation from the statistical part, which largely determines its form.

3. – The Distinction Between Operators and Observables.

The theory of Hilbert space is a necessary part of the mathematical foundation of our theory. The observables of our theory, as in quantum mechanics, are associated with the set of linear Hermitian operators in Hilbert space whose eigenstates form a complete set. In particular, the analysis of such operators

([10]) Ensembles agreeing with quantum mechanics, but having negative measure for some subsets, have been defined. See E. P. WIGNER: *Phys. Rev.*, **40**, 749 (1932); J. P. MOYAL: *Proc. Camb. Phil. Soc.*, **45**, 99 (1949); H. WEYL: *Theory of Groups and Quantum Mechanics* (London, 1931), pp. 274, 275. Such an ensemble cannot have a collection of real physical systems as its counterpart.

given by DIRAC ([11]) will be useful to us as a physically-oriented presentation of their broad features. In this connection, however, a difficulty arises of which we must warn the reader. The closeness of the relationship between the physical observable and its corresponding operator in quantum mechanics has led to a quasi-identification of the two, much used by DIRAC and others, which is useful within the context of quantum-mechanical discussions, but would be misleading for us, because the relationship between them is different in our theory. When a pairwise, one-to-one relationship exists between the members of two sets of quantities, it is often possible to use the same name for both members of a pair without causing any confusion, particularly when the relationship is a very familiar one. But when, as in the present article, different definitions of the relationship are being contrasted, the distinction, namely *operator* vs. *observable*, needs to be carefully maintained.

We should, in fact, warn the reader that in the following text the word « observable » will often not mean what he is used to having it mean. For example, take two operators which do not commute; corresponding to this, one would have the statement in quantum mechanics that their corresponding observables do not both simultaneously « have values » for all physical states in which one of them « has a value ». In quantum-mechanical shorthand, one says that the « observables do not commute ». This is impressionistic language which underlines the parallelism between « observables not simultaneously "having values" » and « operators not commuting ». But in our theory, observables always « have values », not necessarily for the ensemble, but for the individual case. Hence there is no utility for us at present in the *façon de parler* « observables do not commute ».

One note further on terminology: By the phrase « quantum mechanics », whether used substantively, adjectively, or adverbially, we always mean to refer to the standard theory whose clearest and best exposition is probably Dirac's book. It should be observed that from our point of view, however, « quantum mechanics » or « quantum dynamics » are (at least formally) misnomers, since they imply that the systems described are individuals, whereas our position is that they (again, at least formally) really represent ensembles. This is just the dichotomy of viewpoint discussed in § 1. The customary usage is, nevertheless, internally consistent and we think it best to adopt it here. The qualifier « quantum » in these expressions will prevent them from being confused, even in our context, with their true, individual counterparts.

When we refer to « our theory », we shall call it just that, or « the differential-space theory », or « the differential-space theory of quantum systems ».

We now proceed to describe the dynamical specification of individual systems

([11]) P. A. M. DIRAC: *The Principles of Quantum Mechanics*, 3d edition (Oxford, 1947).

in our theory. For simplicity we assume systems that are such that the complete set of independent commuting l.h.c. operators ([12]) required for their observational description in ordinary quantum mechanics ([13]) consists of one operator, i.e., system of « one degree of freedom » quantum-mechanically.

This is not a real restriction. Any system can be reduced to one of a single degree of freedom by the following formal procedure. Consider the product space of the Hilbert spaces of the individual members of its complete set of commuting operators, and take as eigenbasis in the product space the direct products of the eigenvectors of the individual operators. In the product space define an operator, diagonal in this representation, having eigenvalues that may be arbitrary, except that no two shall be equal. This single operator, considered as an observable in the quantum-mechanical sense, is a single degree of freedom that furnishes a representation adequate for the quantum-mechanical description of the system.

Notation. In WS, we used script capitals for observables, bold-face capitals for operators, and ordinary capitals for eigenvalues ($\mathscr{R}$, $\boldsymbol{R}$, and R respectively). Here we discard the use of script capitals in favor of a notation to be given in the third paragraph of the next section. There will be no change in the meaning of bold-face or ordinary capitals.

4. – Dynamical Specification and Dynamics in Differential-Space Theory.

From the set of all non-degenerate l.h.c. operators in the Hilbert space of a dynamical system according to quantum mechanics, to be called N, define a subset I such that no two members of the subset commute, but any member of N not belonging to I commutes with a member of I. Call this subset I. I is in effect a maximal subset of non-commuting non-degenerate operators; it contains one and only one non-degenerate l.h.c. operator for every orientation of principal axes.

Dynamical Postulate 1: *The independent observables of the system are a set of real numerical variables that can be placed in one-to-one correspondence with the members of* I ([14]).

([12]) We use the term « l.h.c. operator » to denote « linear Hermitian operator having a complete set of eigenstates », i.e., an « observable » in Dirac's nomenclature.

([13]) P. A. M. Dirac: loc. cit., § **14**. This set need not consist of constants of the motion.

([14]) This postulate is not intended dogmatically. Ultimately, modifications, refinements, perhaps drastic reductions in the number of independent observables may be possible. The present formulation is aimed at obtaining a dynamics which can be brought into agreement with quantum mechanics through our statistical theory in the easiest possible way.

Notation. The observable corresponding to $\boldsymbol{R}$ in accordance with the above postulate will be symbolized by enclosure in brackets: $[\boldsymbol{R}]$. When an observable is under consideration independently of its relation to any operator, we may use a letter symbol alone, e.g., x or $f(x)$ (cf. next paragraph).

The set I is of course partly arbitrary. Of any subset of N consisting of a given non-degenerate operator and all those that commute with it, only one will be a member of I, but this may be any one. When a choice has to be made in a concrete situation, the variable with the most immediate physical meaning will probably be most convenient, e.g., x (the spatial variable) would be preferable to any $f(x) \neq x$. The choice is in any case not in principle important, since the operator chosen is to be regarded merely as an arbitrary representative of a maximal set of functionally dependent operators, the latter being the important entity ([15]).

As for dependent variables (functions of the independent variables), we first treat functions that do not take on any given value at more than one point of the range of the independent variable. These correspond to non-degenerate operator functions ([16,17]) of the operators of I. The statistical theory to be erected on the dynamical base now being constructed allows us in this case to maintain functional relationships invariant:

Dynamical Postulate 2: *An observable that is a function $f([\boldsymbol{R}])$ of an observable $\boldsymbol{R}$ that does not take on the same value for any two values of $\boldsymbol{R}$, corresponds to the operator $f(\boldsymbol{R})$.*

When this postulate is written down in symbolic form, the invariance property is very clearly displayed:

$$f([\boldsymbol{R}]) = [f(\boldsymbol{R})] .$$

The case of functions that do take on the same value for more than one point of the independent variable requires special treatment, because they are related to degenerate operators. A degenerate operator is a function of (i.e., commutes with) an infinite number of mutually independent non-degenerate operators, one for each distinct basis in the degenerate subspace. The operator $f(\boldsymbol{R})$ alone no longer serves for the correspondence; if the independent operators $\boldsymbol{R}$ and $\boldsymbol{S}$ both commute with $f(\boldsymbol{R})$, $f(\boldsymbol{R})$ is also a function, say $g(\boldsymbol{S})$,

([15]) We might have used, instead of I, the set of distinct eigen-bases of the Hilbert space (spectral families, spectral measures, or resolutions of unity, in the rigorous theory). We prefer to speak in terms of operators rather than eigenbases because the operators are more directly related to the observables of our dynamics.

([16]) P. A. M. Dirac: loc. cit., §§ **11**, **19**.

([17]) B. v. Sz. Nagy: *Spektraldarstellung, linearer Transformationen des Hilbertschen Raumes* (Berlin, 1942; Ann Arbor, 1947)' Chapter X, § 1.

of $\boldsymbol{S}$; but it is necessary for us (as will become clear later, see end of § 5), to maintain a distinction between $f([\boldsymbol{R}])$ and $g([\boldsymbol{S}])$. If we introduce the notion of a degenerate operator $f(\boldsymbol{R})$ *with respect to the basis given by* $\boldsymbol{R}$, i.e., $f(\boldsymbol{R})$ is (arbitrarily) to be assigned this unique eigenbasis only, then $g(\boldsymbol{S})$ becomes in this sense a different operator from $f(\boldsymbol{R})$ and the one-to-one relation between operators and observables can be restored.

Dynamic Postulate 2′: *An observable that is a function $f([\boldsymbol{R}])$ of an observable $[\boldsymbol{R}]$, and that does not take on different values for all points of the range of $[\boldsymbol{R}]$, corresponds to the operator $f(\boldsymbol{R})$ taken with respect to the basis given by $\boldsymbol{R}$.*

We now have a complete correspondence for the dynamical variables of a system, in the sense that all l.h.c. operators have variables assigned to them, and there are no other variables (except in the sense that we have gone somewhat beyond the usual definition of an l.h.c. operator in the case of degenerate operators). This set of variables is the minimum that is necessary and sufficient for the statistical correspondence we later set up with quantum mechanics.

Our next postulate defines the possible values of an observable and takes the simple — and familiar — form:

Dynamic Postulate 3: *The only possible values of an observable $[\boldsymbol{R}]$ are the eigenvalues of the operator $\boldsymbol{R}$.* (This is to hold whether $\boldsymbol{R}$ belongs to I or not, i.e., $\boldsymbol{R}$ is any l.h.c. operator). Thus the quantization of observed values is not « explained » in our theory any more than in quantum mechanics, but arises in the same way from a postulated correspondence with l.h.c. operators.

The postulates up to now have been concerned only with the dynamic specification of systems, i.e., with the framework of variables to be used in the description of the evolution of the system. The remaining discussion of this section deals with the dynamics proper, i.e., with the time-evolution of the system.

First we need Dynamic Postulate 4: *Associated with any closed system of given physical properties, there is a unitary operator $\boldsymbol{U}(t)$, depending on the time t in such a way that, for arbitrary t', t''*

$$\boldsymbol{U}(t')\boldsymbol{U}(t'') = \boldsymbol{U}(t' + t'') \,. \tag{1}$$

It will not startle the reader if we prematurely reveal that $\boldsymbol{U}(t)$ is the same as the time-displacement operator one would use for the system in quantum mechanics. It should nonetheless be appreciated that the precedence of the quantum-mechanical time-displacement operator is historical only, and that —within the universe of discourse of the individual dynamics under discussion — the rôle played by $\boldsymbol{U}(t)$ is strictly with respect to this individual dynamics

and does not require the concept of the wave function, i.e., of the quantum ensemble considered as a whole, to give it meaning.

We are now in a position to state the basic dynamic principle. Let $[\boldsymbol{R}](t)$ be the value of the observable $[\boldsymbol{R}]$ at the arbitrary time t. Then we have the single equation that is DYNAMIC POSTULATE 5:

$$[\boldsymbol{R}](t) = [\boldsymbol{U}^{-1}(\tau)\boldsymbol{R}\boldsymbol{U}(\tau)]_{\tau=t}(0) \,. \tag{2}$$

It would have been perfectly correct to have written simply $[\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)](0)$ for the right-hand side of (1), but the somewhat awkward notation here employed is temporarily necessary to make the meaning unmistakable. $[\boldsymbol{U}^{-1}(\tau)R\,U(\tau)]_{\tau=t}(T)$ is an observable that is itself a function of time, but its time variable is T; so far as this observable is concerned, t is not its time variable but a parameter that defines it, i.e., singles it out from the other observables $[\boldsymbol{U}^{-1}(\tau)\boldsymbol{R}\boldsymbol{U}(\tau)](T)$, functions of time T, but with values of τ other than t.

Having established our meaning, we shall henceforth use, instead of (2), the more succinct form

$$[\boldsymbol{R}](t) = [\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)](0) \,. \tag{2'}$$

Equations (2) and (2′) were written with respect to a fixed origin of time. They imply, however, equations of the same form for any choice of the time origin. From (2′) we have, with $t = t_0 + t'$,

$$\begin{aligned}[\boldsymbol{R}](t_0 + t') &= [\boldsymbol{U}^{-1}(t_0 + t')\boldsymbol{R}\boldsymbol{U}(t_0 + t')](0) \\ &= [\boldsymbol{U}^{-1}(t_0)\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')\,\boldsymbol{U}(t_0)](0)\end{aligned}$$

(from equation (1)):

$$= [\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')](t_0) \tag{3}$$

(from equation (2′), with $\boldsymbol{R}$ replaced by $\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')$ and t replaced by t_0). Displacement of the time-origin to the right by an amount t_0 gives a new function $[\boldsymbol{R}]'(t')$ of $t' = t - t_0$, equal to $[\boldsymbol{R}](t)$. Thus (3) becomes

$$[\boldsymbol{R}]'(t') = [\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')]'(0) \,, \tag{4}$$

identical in form with (2′).

Equation (3), i.e.,

$$[\boldsymbol{R}](t_0 + t') = [\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')](t_0) \,, \tag{5}$$

considered for a variable t_0 and a constant t', reveals that the observables $[\boldsymbol{R}](t)$ and $[\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')](t)$ differ from one another only through a time displacement, i.e., at a given instant $[R]$ has the value $[\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')]$ had at an instant earlier by an amount t'. In this respect consider a situation in which $\boldsymbol{R}$ and $\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')$ do not commute. Then both may be chosen as distinct members of I, and $\boldsymbol{R}$ and $\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')$ are therefore, by Dynamical Postulate 1, independent variables. The reasoning following equation (5) shows, however, that this is true only if both are considered at the same instant of time. $[\boldsymbol{R}]$ at a time t' later than t_0 *is* dependent on, in fact equal to, $[\boldsymbol{U}^{-1}(t')\boldsymbol{R}\boldsymbol{U}(t')]$ at t_0. All this, however, does not invalidate but only makes precise the independence of the observables $[\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)](t_0)$ with various values of t. Holding t_0 fixed, and assuming the past and future (relative to t_0) absolutely unspecified, the assignment of arbitrary (within our previously defined limitations), mutually independent values to the set of observables obtained by letting t take on all values from $-\infty$ to $+\infty$, results in a determined past and future for these variables; but this agrees with any intuitive concept of a complete set of independent observables for a sharply defined dynamic system. (Note, however, that any observable whose corresponding operator is not equal to $\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)$ for any value of t is undetermined by the past values of all these operators $\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)$, thus we see that the values of all these operators for a given instant do not determine all the variables of the system; observables of a system will fall into disjoint sets such that observables are dynamically related if and only if they belong to the same set.)

There is a similarity between the dynamical principle here stated and classical dynamics in one of its formulations. It is well known that ([18]) in classical systems there exists a certain contact transformation, continuous function of a parameter (say, λ), of the present variables of a system, such that the transformed variable for a given value of λ at zero time is equal to the value of the untransformed variable at time $t=\lambda$. The parallel with our dynamics is obvious. There is, however, a difference. In the classical theory, one has a certain set of independent variables as untransformed variables, and the transformation involves rather general functional relationships between these and the transformed variables. In our theory, one has again a set of independent variables, but the transformation is of a simpler type, the transformed set being *as a set* identical with the untransformed: that is, the transformation is just a re-labelling of the original variables.

It is readily apparent that the Dynamical Postulates above stated define a system that satisfies condition 1 of § 2.

We note one important property of the dynamics here defined: Observables whose operators commute with the Hamiltonian, and therefore with

([18]) H. GOLDSTEIN: *Classical Mechanics* (Cambridge, Mass., 1950), § 8-6.

$\boldsymbol{U}(t)$, are *constants of the motion* of our individual systems, since if $\boldsymbol{R}$ commutes with $\boldsymbol{U}(t)$, $\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)=\boldsymbol{R}$ and eq. (2′) gives

$$[\boldsymbol{R}](t) = [\boldsymbol{R}](0) .$$

Thus observables that are « constants of the motion » in the somewhat special sense of quantum mechanics are quite literally constants of the motion in our theory.

5. – Summary of Properties of Differential Space; Restatement of the Basic Theorem of Differential-Space Quantum Theory.

In this section we set the stage for the presentation of the statistical part of our theory. There will be a brief, selective summary of the material in WS that is relevant to the present purpose, and an improved proof of the theorem in that paper, the « polichotomic theorem », that is basic here. We shall also, later in the section, find it appropriate to introduce an important expository simplification concerning operators having a continuous spectrum. Finally, we supplement the work of WS by showing how the polychotomic theorem is to be applied in the case of a degenerate operator.

First, we briefly characterize differential space ([19]) in the form it takes with respect to a *discrete basis*: It is an extended Hilbert space, (1) including all un-normalizable vectors as well as the normalizable vectors to which Hilbert space is restricted; and (2) in it is defined a measure associated with a given volume differential as follows (we assume at first a discrete spectrum, and discuss continuous spectrum operators later). Given a non-degenerate l.h.c. operator $\boldsymbol{R}$ and its associated eigenbasis consisting of the set of orthonormal eigenkets $|R_i\rangle$, each being labelled by its associated eigenvalues R_i ([20]). A

([19]) N. WIENER: *Journ. Math. Phys.*, **2**, 131-174 (1923). An extended heuristic account of differential space is contained in WS, and here will also be found further references on the subject.

([20]) The introduction of the l.h.c. operator $\boldsymbol{R}$ here is not strictly necessary for the immediate argument. The basis is entirely arbitrary, since an operator can always be constructed whose eigenkets are exactly the basic kets of a previously assigned basis. The operator $\boldsymbol{R}$ is introduced only in order to unify the present stage of the argument with later work.

point P ([21]) in differential space is characterized by a vector whose components with respect to these axes are, say, $\xi_i(P)$. More precisely: If the vector to P has the ket $|P\rangle$, $\xi_i(P)$ is the component of the *bra* $\langle P|$ with respect to the eigenaxis belonging to R_i, i.e.,

$$\xi_i(P) = \langle PR_i \rangle \,. \tag{5a}$$

This definition insures agreement with the formal work of WS. ξ_i is complex; separating real and imaginary parts,

$$\xi_i = \xi_{iR} + i\xi_{iI} \,. \tag{6}$$

A weight or probability ([22]) is assigned to the set of points P contained in a small volume element

$$\prod_i \mathrm{d}\xi_{iR}\,\mathrm{d}\xi_{iI} \tag{7}$$

at the point P', equal to

$$\prod_i (2\pi)^{-1} \exp[-|\xi_i|^2/2]\,\mathrm{d}\xi_{iR}\,\mathrm{d}\xi_{iI} = \prod_i (2\pi)^{-1} \exp[-x_i]\,\mathrm{d}x_i\,\mathrm{d}\theta_i; \tag{8}$$

θ_i and x_i are the phase and half the squared modulus of ξ_i respectively:

$$\xi_i = \sqrt{2x_i}\,\exp[i\theta_i] \,. \tag{9}$$

The increments $\mathrm{d}\xi_{iR}$ and $\mathrm{d}\xi_{iI}$ are always taken positive, i.e., the element of probability is necessarily non-negative.

Two properties of the above defined weight are basic for the theory of quantum systems.

([21]) In WS and in WIENER's work on differential space (v. foot-note ([19])) the « parameter of distribution » α identifies differential space points and also acts as the variable of a space in which the weight of eq. (8) is expressed as a direct measure of sets of points. P is, with some extra emphasis on the vector interpretation, essentially equivalent to α in its first function, i.e., relieved of the burden of bearing the weight expressed by eq. (8).

([22]) The *mean square* of any ξ_{iR} or ξ_{iI} in the distribution (8) is unity. This makes the mean over P of $|\xi_i(P)|^2$ or the absolute square of the projection, on to *any* normalized vector, of the vector to P, equal to 2. We take this occasion to point out an error in § 4 of WS. If eq. (19) there is accepted, $\int |\psi(\alpha)|^2 \mathrm{d}\alpha = 2$, not 1, and corresponding changes need to be made in eq. (22), the equation following (23), and eq. (24). When $\int |\psi(\alpha)|^2 \mathrm{d}\alpha \neq 1$ the transformation from x to α (WS end of § 2) is not strictly unitary. It can, of course, be made so by making the mean squares of $|\xi_{iR}|$ and $|\xi_{iI}|$ equal to $\frac{1}{2}$ each.

(1) *Unitary invariance.* If all the points P of a set S are transformed by a unitary operator (realized by a unitary matrix acting on the ξ_i) into points P' defining a set S', the weight of S' equals that of S. This may be understood as a hyperspherical symmetry of the weight function in differential space.

(2) *Statistical independence of the* ξ_i. The weight is a product of factors each containing only one ξ_i. This, combined with unitary invariance, means that the components of a vector in differential space with respect to *any* two orthogonal unit vectors are distributed independently of one another. This property is especially important for the continuous spectrum. The Gaussian form of distribution used here is the only form possessing both of these properties, and having finite mean squares for all variables.

The basic theorem that will be used in constructing the statistical part of the present work is that of the *polychotomic method* described in § 5 of WS, which we summarize here with a shortened version of the proof given there in eqs. (35)-(38). Let there be given a normalized quantum-mechanical state vector whose representative in the $\boldsymbol{R}$ basis is

$$a_i = \langle R_i | \psi \rangle . \tag{10}$$

We now define a certain correspondence depending on ψ between eigenvalues R_i and points P, i.e., we define a functional $R(P, \psi)$ which may take on only the values R_i. (Although the relation between P and ψ in this work is not without a certain symmetry, in a simple interpretation ψ is to be considered as fixed and P as variable). $R(P, \psi)$ is to be, by definition, that eigenvalue R_k for which the ratio of the magnitudes of the associated quantities $\xi_k(P)$ and a_k is a minimum, i.e.,

$$R(P, \psi) = R_k , \tag{11}$$

if and only if

$$\frac{\xi_k(P)}{|a_k|^2} < \frac{\xi_i(P)}{|a_i|^2} \qquad \text{(all } i\text{)}. \tag{12}$$

We shall refer to this method of constructing $R(P, \psi)$ as the *polychotomic algorithm.*

We now show that for any given k the weight of the region satisfying (12), obtained as the integral of the elementary weight expression (8), over this region, is just $|a_k|^2$. In symbols:

$$\text{Prob. } \{P \text{ satisfies (12)}\} = \int \cdots \int \prod_i (2\pi)^{-1} \exp[-x_i] \, \mathrm{d}\theta_i \, \mathrm{d}x_i = |a_k|^2 , \tag{13}$$

the integral being extended over the region satisfying (12). (We shall call this the *polychotomic theorem*). Eq. (13) is most easily proved if there are only a finite number, say n, of non-vanishing a_i; one can then apply the result to any wave function by going to the limit $n \to \infty$ if necessary. Imagine, for ease in writing, the a_i renumbered so that a_k becomes a_1 and the remaining $n-1$ non-vanishing a_i receive subscripts from 2 to n inclusive. Then (12) reads (changing to the x_i as variables)

$$\frac{x_1}{|a_1|^2} < \frac{x_i}{|a_i|^2} \qquad \text{(all } i\text{)}. \tag{12'}$$

All ξ_i are to be treated as independent variables of integration subject only to the inequalities (12'). These imply no restriction on the θ_i, so the angle integrations (from 0 to 2π) of (13) may be done at once, yielding

$$\int \cdots \int \prod_i \exp[-x_i]\,\mathrm{d}x_i\,. \tag{14}$$

The inequalities (12') imply limits of integration on the individual x_i independent of one another, as follows:

$$x_i > \frac{|a_i|^2}{|a_1|^2} x_1 \qquad (i = 2, 3, \ldots, n), \tag{15a}$$

$$\infty > x_1 > 0\,, \tag{15b}$$

$$\infty > x_i > 0\,, \qquad (i > n). \tag{15c}$$

For $i > n$, each x_i may be integrated independently, giving a factor 1. We remain with

$$\int_0^\infty \exp[-x_1]\,\mathrm{d}x_1 \int_{\frac{|a_2|^2}{|a_1|^2}x_1}^\infty \exp[-x_2]\,\mathrm{d}x_2 \int_{\frac{|a_3|^2}{|a_1|^2}x_1}^\infty \exp[-x_3]\,\mathrm{d}x_3 \cdots \int_{\frac{|a_n|^2}{|a_1|^2}x_1}^\infty \exp[-x_n]\,\mathrm{d}x_n = |a_1|^2. \tag{16}$$

(remembering that $\sum |a_i|^2 = 1$ due to normalization). This (since $|a_1|^2$ is equal to what in the original numbering was $|a_k|^2$) proves the theorem.

The above proof of the polychotomic theorem is equivalent to the work of eqs. (35) to (38) of WS.

If part or all of the spectrum of R is *continuous*, then in a formal sense one thinks of the discrete index i of $\xi_i(P)$ as replaced by the continuous variable R to give, say, $\xi(R; P)$. But this transition cannot be so simply carried out in the distribution (8). $\xi(R; P)$ is a component of the vector to P only

in an improper sense. The distribution (8) is defined only for resolutions of the vector to P with respect to a discrete basis. Various devices may be employed to adapt the polychotomic algorithm to the continuous case, but all have the property of introducing discrete bases. One is given in eqs. (40) and (41), plus the material in between, of WS; another, having the advantage of making the base vectors that are introduced independent of the wave function, would be to use eigendifferentials.

Although the inclusion of the continuous spectrum does not introduce any difficulties into the polychotomic algorithm, it does introduce complications of a purely formal character into the theory being presented here. For the purposes of the present discussion, therefore, we consider it best to adopt the tentative fiction that all operators have discrete spectra only. This will allow the essentially physical line of argument here being presented to emerge with much greater clarity. The *physical picture* is not different for the continuous spectrum. Thus, having presented it here in unified fashion, it will be easy in later work to show how the continuous spectrum is incorporated into it.

The statement of the polychotomic algorithm given above, and the work of WS, are adequate only if the operator $\boldsymbol{R}$ is non-degenerate. Consider now a degenerate operator, which we call $\boldsymbol{D}$ for the sake of distinguishing it from non-degenerate operators. $\boldsymbol{D}$'s eigenkets $D_{i,j}$ are not unique, and this makes the value of $D(P, \psi)$ non-unique over a set of non-zero measure (this assertion is proved in the Appendix). Some way of fixing the eigenkets of $\boldsymbol{D}$ is needed in order to make $D(P, \psi)$ unique, or, rather, to give unique labels to the infinite number of different functionals $D(P, \psi)$. In order to do this in a physical way, we note that each member of the infinite set of different complete orthogonal bases of $\boldsymbol{D}$ is the unique eigenbasis of some l.h.c. operator belonging to the set I (§ 4). Select any pair, say $\boldsymbol{R}$ and $\boldsymbol{S}$ of such members of I whose unique eigenbases are eigenbases of $\boldsymbol{D}$. Then $\boldsymbol{R}$ and $\boldsymbol{S}$ commute with $\boldsymbol{D}$ but not with one another, and one has $\boldsymbol{D}$ a function of either:

$$\boldsymbol{D} = f(\boldsymbol{R}) = g(\boldsymbol{S}) \,. \tag{17}$$

If the eigenbasis of $\boldsymbol{R}$ is used in applying the polychotomic algorithm to $\boldsymbol{D}$, one obtains uniquely $\boldsymbol{D}(P, \psi) = f(R(P, \psi))$; if that of $\boldsymbol{S}$ is used, one obtains, again uniquely, $D(P, \psi) = g(S(P, \psi))$. Thus a unique selection of a functional $D(P, \psi)$ is obtained by specifying along with $\boldsymbol{D}$ a non-degenerate operator with which it commutes. We can absorb the degenerate case into the terminology of the non-degenerate case if we agree to consider $f(\boldsymbol{R})$ and $g(\boldsymbol{S})$, whenever explicitly so written, as having distinct eigenbases, viz.: $f(\boldsymbol{R})$'s eigenbasis is that of $\boldsymbol{R}$, and $g(\boldsymbol{S})$'s, is that of $\boldsymbol{S}$. The inexplicit notation $\boldsymbol{D}$ is rejected, under this agreement, as not adapted for the purpose of the differential-space theory. An operator is defined by its eigenbasis, not merely by its effect on a vector.

In the next section we explicitly introduce the notion of $R(P, \psi)$'s ($\boldsymbol{R}$ any operator) being a value of an observable belonging to a system associated with the point P. The preceding paragraph is then seen to give the statistical background of Dynamical Postulate 2′ and the paragraph preceding it, § 4.

6. – Statistical Part of the Theory.

Given a system whose observables in quantum mechanics are such as to correspond to the set of all l.h.c. operators in a certain Hilbert space. Then, since $\boldsymbol{R}$ in the polychotomic algorithm is any l.h.c. operator, we can assign a value of each and every observable (via its associated l.h.c. operator) [23] to any point P in differential space, except for a set of measure zero. We take the set of all points of differential space to be the mathematical representation of an ensemble [24] of systems, each point [25] corresponding to a system the values of whose observables are those associated with the point. For the measure of any suitably chosen [26] subensemble of the ensemble we take the differential-space measure (8), integrated over the associated subset of differential space. We thus have an ensemble of systems that are sharply defined in Einstein's sense, namely that of the simultaneous assignment to the system of precise values of all physical observables, with a positive measure for all subensemble, and such that every statistical property derivable from the wave function by quantum mechanics is verified for the ensemble (the ensemble is, of course, richer in statistical properties than the quantum-mechanical algorithm). Hence we have, so far as condition 2 of § 2 is con-

[23] Strictly speaking, for the continuous spectrum we can assign only an arbitrarily small interval of eigenvalues, not an individual value. This « hedge » illustrates the trivial logical difficulties (mathematically, of course, they are not trivial) introduced into the exposition by the continuous spectrum, and helps to show why we chose to omit it for the present.

[24] We use « ensemble » in Gibbs's sense, i.e., it is a collection of objects used to form a statistical representation of a physical situation in which one's prior knowledge is less than maximal. The word « set » is used in the mathematical sense, e.g., « the theory of sets. » Any ensemble is, of course, also a set, but we prefer to be more specific when we actually refer to an ensemble. We use the word « ensemble » as if it were English; the fact that « ensemble » means « set » in French does not make it mean « set » when it is used as an English word.

[25] This is of course somewhat loosely stated. The points of differential space are a continuum, and cannot be put in one-to-one correspondence with any set of discrete objects. This correspondence is more rigorously obtained, as in the phase space of classical statistical mechanics, by associating small regions of finite measure, which can be shrunk down to infinitesimal size in the limit, with systems.

[26] The corresponding subset of differential space must be Borel measurable. This is not an onerous requirement, physically speaking.

cerned, carried out our program of a representation of quantum-mechanical statistical properties through ensembles of sharply-defined systems. This means, essentially, that so far we can establish a suitable ensemble at any instant. We have yet to satisfy conditions 1 and 3, which have to do with the temporal changes of individual systems and ensembles respectively.

Condition 3 is now readily satisfied by having all representative points move in differential space according to some unitary transformation. This alone, according to the unitary invariance property of the weighting (8), is enough for the purpose. Thus we still have a free choice as to the particular unitary transformation we wish to use.

The most natural choice for this unitary transformation is the Schrödinger time-displacement operator $\boldsymbol{U}(t)$. That is, the ket at t for a system whose ket at 0 was $|P\rangle$, is $\boldsymbol{U}(t)|P\rangle$. (We shall also say that P is carried into UP, even though the transformations of P and of $|P\rangle$ are formally different sorts of things). In effect, $|P\rangle$ satisfies the same Schrödinger equation as the state vector $|\psi\rangle$ ([27]).

The most important consequence of this choice is the simple deterministic form it gives for the individual dynamics of systems. That is, condition 1 of § 2 is satisfied. To prove this we need the following lemma:

$$R(\boldsymbol{W}P, \boldsymbol{W}\psi) = (\boldsymbol{W}^{-1}R\boldsymbol{W})(P, \psi)\,, \quad ([28]) \tag{18}$$

if $\boldsymbol{W}$ is a unitary transformation. Equation (18) is proved as follows: Let

$$R(\boldsymbol{W}P, \boldsymbol{W}\psi) = R_k\,. \tag{19}$$

This is equivalent to

$$\frac{|\langle R_k|\boldsymbol{W}P\rangle|}{|\langle R_k|\boldsymbol{W}\psi\rangle|} < \frac{|\langle R_m|\boldsymbol{W}P\rangle|}{|\langle R_m|\boldsymbol{W}\psi\rangle|}\,, \qquad \text{all } m, \tag{20}$$

But since $|WP\rangle = W|P\rangle$, $|W\psi\rangle = W|\psi\rangle$, we have

$$\frac{|\langle R_k|\boldsymbol{W}|P\rangle|}{|\langle R_k|\boldsymbol{W}|\psi\rangle|} < \frac{|\langle R_m|\boldsymbol{W}|P\rangle|}{|\langle R_m|\boldsymbol{W}|\psi\rangle|}\,, \qquad \text{all } m. \tag{21}$$

([27]) Since the space representative of $|P\rangle$ is, typically, not a smooth function, there is no question of its satisfying the Schrödinger equation in its differential form. Either *a*) the integral form must be used, *b*) an expansion in energy eigenfunctions is carried out, with each individual eigenfunction propagating itself in time according to the factor $\exp[-iEt/\hbar]$, or *c*) the function is considered as the limit of a sequence of functions that do satisfy the Schrödinger equation, and its motion in differential space is then obtained as that of the limit function of this sequence. All three methods, of course, give the same result.

([28]) The right-hand side is to be read as if it were $f(P, \psi)$, with $f = (\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})$.

Now

$$R\langle_k|\,\boldsymbol{W} = \langle(\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})_k|\,, \tag{22}$$

provided we number alike equal eigenvalues of $\boldsymbol{R}$ and of $\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W}$. Thus (21) may be written

$$\frac{|\langle(\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})_k|P\rangle|}{|\langle(\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})_k|\psi\rangle|} < \frac{|\langle(\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})_m|P\rangle|}{|\langle(\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})_m|\psi\rangle|}\,, \qquad \text{all } m. \tag{23}$$

which is equivalent to

$$(\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})(P, \psi) = R_k\,. \tag{24}$$

Comparing this with (19), one obtains equation (20).

Eq. (20) may be « proved » geometrically as follows:

The two orthogonal axes in Fig. 1 represent the entire infinite-dimensional set of axes $|R_i\rangle\cdot|R_k\rangle$ is singled out and labelled. The unitary transformation $\boldsymbol{W}$ is represented as a rotation. The vectors $|P\rangle$ and $|\psi\rangle$, and the rotated vectors $W|P\rangle$ and $W|\psi\rangle$ are shown. It is apparent that $|P\rangle$, $|\psi\rangle$ are to the inversely rotated axes $\boldsymbol{W}^{-1}|R_i\rangle = |(\boldsymbol{W}^{-1}\boldsymbol{R}\boldsymbol{W})_i\rangle$, exactly as $\boldsymbol{W}|P\rangle$, $\boldsymbol{W}|\psi\rangle$ are to the axes $R_i\rangle$. This gives (18) immediately.

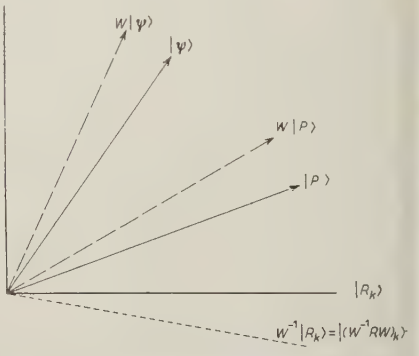


Fig. 1. - Schematic proof of Eq. (18).

We can now show easily that the time evolution of systems in the ensemble is given by exactly the dynamics of Dynamic Postulate 5, § **4**. The definition given of the motion of the system point with time means that $[\boldsymbol{R}](t)$ for the system represented by P at $t = 0$ satisfies

$$[\boldsymbol{R}](t) = R(\boldsymbol{U}(t)P, \boldsymbol{U}(t)\psi) \tag{25}$$

for all t. But eq. (18), with $\boldsymbol{W} = \boldsymbol{U}(t)$, gives then

$$[\boldsymbol{R}](t) = (\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t))(P, \psi)\,. \tag{26}$$

If (25) is written down for an arbitrary operator $\boldsymbol{S}$

$$[\boldsymbol{S}](\tau) = S(U(\tau)P,\ U(\tau)\psi)\,, \tag{27}$$

then putting $\tau = 0$ gives

$$[S](0) = S(P, \psi)\,, \tag{28}$$

hence the right-hand side of (26) is equal to $[\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)](0)$ for the same system, and we have

$$[\boldsymbol{R}](t) = [\boldsymbol{U}^{-1}(t)\boldsymbol{R}\boldsymbol{U}(t)](0)\,, \tag{29}$$

the same as eq., (2′), i.e., Dynamic Postulate 5.

The fact that (29) applies to the systems of our ensemble means that the specification for a system, at a given time, of the values of the observables $[\boldsymbol{U}^{-1}(\tau)\boldsymbol{R}\boldsymbol{U}(\tau)]$ for all $\tau > t$ suffices to determine its future, and therefore condition 1 of § 2 is satisfied. In particular, a knowledge of ψ is unnecessary in predicting the evolution of a system if these observables are known, as it should be if the system is a single deterministic system.

Since we have now satisfied conditions 1, 2 and 3 of § 2, we have demonstrated that quantum mechanics is equivalent to a theory of crypto-deterministic processes, if these conditions are considered acceptable criteria. In this connection one objection might be anticipated: One might argue that a description containing an infinite number of parameters is not deterministic, operationally speaking, since it is not experimentally realizable in a finite time. Against this argument we would assert that it would also make classical mechanics operationally indeterministic, since perfete precision in even only one measurement would require an infinite amount of time; and it would certainly make classical wave theory indeterministic, since the specification of a field requires its measurement at all points of a continuum.

It will be noted that, in satisfying condition 1 of § 2, we have appealed only to Dynamic Postulate 5 of § 4. This is the only one of the postulates necessary in satisfying this condition. We should therefore show why the other postulates are given at all. Generally speaking, they are required for completeness in the description of the individual-particle dynamics as such. More specifically: In § 4, a correspondence is set up between $\boldsymbol{Q}$ and the *observable* $[\boldsymbol{Q}]$; the polychotomic algorithm sets up a correspondence between $\boldsymbol{Q}$ and the *functional* $Q(P, \psi)$. We have a crypto-deterministic theory of quantum systems once we set up a correspondence between the *observable* $[\boldsymbol{Q}]$ and the *functional* (until now of merely mathematical status) $Q(P, \psi)$. But the polychotomic algorithm, and the discussion in the next-to-last paragraph of § 5, also yield statements about the correspondence $\boldsymbol{Q} \rightarrow Q(P, \psi)$ when $\boldsymbol{Q}$ is regarded as a function of an operator other than $\boldsymbol{Q}$. Thus the correspondence $\boldsymbol{Q} \rightarrow |\boldsymbol{Q}|$ may not be set up arbitrarily in this case, but rather defined so that $|\boldsymbol{Q}|$ has

the correct relationship to the observable corresponding to the operator of which $\boldsymbol{Q}$ is a function. This is the purpose of postulates 2 and 2′.

The necessity of postulates 3 and 4, in adapting the dynamics to the eventual statistical theory, is, we think, self-evident.

Postulate 1 is fundamental, in that it defines the correspondence $\boldsymbol{Q} \rightarrow [\boldsymbol{Q}]$, and this is its primary *raison d'être*. That the statistical theory does indeed allow the observables associated with the members of I to be independent (i.e., that it generates no functional relationship among them) will not be proved here, but left for a later paper.

At this point we mention again that, as pointed out at the end of § 5, by satisfying eq. (2′), the dynamics of the individual systems of our ensemble makes observables that are « constants of the motion » in quantum mechanics, literally constant for these individual systems. Thus all the conservation laws of a quantum mechanical system, and also of the classical system having the same formal Hamiltonian, are satisfied by them.

A later article will deal with possible physical interpretations of the theory, with its relationship to other hidden-parameter theories (especially that of Bohm), with possible variations in the formulation, and other matters.

* * *

We wish to thank Dr. Saul T. Epstein for many interesting discussions, which contributed to a more clean-cut understanding of the physical interpretation of the quantum-mechanical wave function.

Appendix

Proof that the Result of the Polychotomic Algorithm is Non-Unique When Applied to a Degenerate Operator.

We here show (see § 5) that, when $\boldsymbol{D}$ is degenerate, the value of $D(P, \psi)$ is non-unique over a set of non-zero measure if only the unaltered polychotomic algorithm is used.

Suppose, at some point P, $D(P, \psi) = D_k$ with D_k degenerate. Then (see eq. (14)), $\xi_k(P)$, the absolute magnitude of whose ratio to a_k is the smallest of all such ratios at P, is the component of $|P\rangle$ with respect to only one of a degenerate set of, say, ν eigenkets $|D_k 1\rangle$, $|D_k 2\rangle$, ... $|D_k \nu\rangle$, all having eigenvalue D_k, and should be distinguished from the others — let us do so by calling it number 1 of the sequence, $|D_k 1\rangle$. Now a unitary transformation that affects only this degenerate set of eigenkets changes them into a new set of eigenkets, say $|D_k 1'\rangle$, $|D_k 2'\rangle$, ... $|D_k \nu'\rangle$ ($\nu' = \nu$). Let the original ξ's be

denoted by $\xi_{k1}, \xi_{k2}, \ldots, \xi_{k\nu}$ and the original a's by $a_{k1}, a_{k2}, \ldots, a_{k\nu}$; the transformed ξ's and a's, by the same symbols primed. Now it is certainly possible for the P under consideration to be a point such that of the new ratios in the degenerate space $|\xi'_{k\alpha}|/|a'_{k\alpha}|$ $(\alpha = 1, 2, \ldots, \nu)$ none is a minimum in the entire Hilbert space even though $|\xi_{k1}|/|a_{k1}|$ was such a minimum. We set up an extreme situation for the sake of a more striking illustration of this possibility: choose P so that some $|\xi_{ki}|/|a_{ki}|$ for which $D_i \neq D_k$, say $|\xi_1|/|a_1|$ is just barely greater than the minimum $|\xi_{k1}|/|a_{k1}|$.

Let $\nu = 2$ for simplicity. Then there is only one other eigenket with $D = D_k$, namely $|D_k 2\rangle$. Choose $|\xi_{k2}|$, which is still at our disposal, so that $|\xi_{k2}|/|a_{k2}| \gg |\xi_{k1}|/|a_{k1}|$ and $|\xi_{k2}| \gg |\xi_{k1}|$. Take a transformation of the subspace of $|D_{k1}\rangle$ and $|D_{k2}\rangle$ that mixes them in roughly equal amounts. Then, due to the second of the preceding inequalities, $|\xi'_{k1}|$ and $|\xi'_{k2}|$ will each be roughly equal to $\frac{1}{2}|\xi_{k2}|$, which is much greater than $|\xi_{k1}|$. This fact, with the rest of the two preceding inequalities, shows that $|\xi'_{k1}|/|a'_{k1}| \gg = \xi_{k1}|/|a_{k1}|$ and $|\xi'_{k2}|/|a'_{k2}| \gg |\xi_{k1}|/|a_{k1}|$. Thus, since $|\xi'_1|/|a'_1| = |\xi_1|/|a_1|$ (the transformation affects only eigenkets for $R = R_k$), and $|\xi_1|/|a_1|$ is almost as small as $|\xi_{k1}|/|a_{k1}|$, $|\xi'_1|/|a'_1|$ is the smallest of all the ratios after the transformation.

To see that the set in differential space for which the above can hold is of non-zero measure, one need only observe that non-zero variations of all the $|\xi_i|$ about the particular point P are possible without affecting the situation. In fact, the range of variation simply increases with the strength of the inequalities of the preceding paragraph.

Obviously the reverse of the above example is possible, i.e., there are sets of non-zero measure for which $D(P, \psi)$ changes from some arbitrary value into a degenerate eigenvalue. The non-uniqueness can be generally characterized as follows: the transformations that interchange complete orthonormal sets of eigenkets of a degenerate l.h.c. operator will (except for the identity transformation) transform the functional $D(P, \psi)$ over a set of P of non-zero measure. At a given point P, at least one of the two values of D that are interchanged must be a degenerate eigenvalue.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

The General Statistical Problem in Physics and the Theory of Probability.

D. Bohm (*) and W. Schützer

Faculdade de Filosofia, Ciências e Letras - Universidade de São Paulo

(ricevuto il 15 Marzo 1955)

Contents. 1. Introduction. – 2. Chance and Statistical Law. – 3. The Concept of Probability. – 4. The Irregular Collectives of von Mises. – 5. The Modern Axiomatic School. – 6. On the General Statistical Problem as It Arises in Statistical Mechanics. – 7. A Further Example of the General Statistical Problem. The Deduction of a Probabilistic Law from a Determinate Law. – 8. The General Applicability of the Calculus of Probability. – 9. A Criterion for Randomness. – 19. Further Illustration in Terms of Random Phases. – 11. Conclusion.

1. – Introduction.

Basic statistical problems in physics, (e.g., those connected with statistical mechanics and the quantum theory) have, as in other fields, so generally been given a mathematical formulation in terms of the theory of probability, that the impression is created that perhaps no other kind of formulation of such problems is even possible. In the present paper, however, we wish to raise the question of whether the theory of probability really represents the most adequate possible mathematical instrument in all of those cases to which it is customarily applied. Indeed, as we shall see in this paper, there is a wide range of statistical applications in physics, in which a more reliable and a clearer treatment can be obtained by using methods other than those of the theory

(*) Now on leave of absence at Technion, Haifa Institute of Technology, Haifa, Israel.

of probability. In the study of such applications, one sees that what is relevant is not so much the theory of probability, but rather what we call « the general statistical problem », of which the problem of probability is only a special case.

The general statistical problem may be stated thus: To find those relationships which hold in a statistical aggregate independently of a wide range of variations of the detailed properties of the individual members of the aggregate, the size of this range of variation depending on the physical conditions and the context of interest.

It is evident that if relationships of the type described above hold, then whatever may be the character of these relationships, we shall be able to make statistical predictions concerning the aggregate without having to treat the problem of just what the individual members are doing. In those cases to which the theory of probability applies, such statistical predictions can be made with the aid of the well known probability calculus; but as we shall show in this paper, there are a number of rather simple problems, in which it is highly advantageous to go outside the scheme of expressing the statistical relationship in terms of nothing but the concept of probability. These examples will also provide further insight into the meaning of the concept of probability; for they enable us to see this concept inside of the broader setting of statistical laws in general.

After formulating the general statistical problem with the aid of certain examples drawn from the field of statistical mechanics, we shall focus our attention on a few cases to which the theory of probability applies. Here, we shall discuss the problem of how probabilities are related to determinate laws, and even to a more general class of laws lying between the determinate and the probabilistic types. Such a process of relating probabilistic laws to other types of laws will also enable us to see more clearly what is the nature of randomness. Moreover, as we shall see, studies of this kind lead us into new kinds of problems that cannot even be formulated in terms of the theory of probability itself, such as that of estimating in specific cases the error and domain of applicability of various aspects of the theory of probability.

The general plan followed in this paper will be first to review some of the basic quantitative ideas on the theory of probability that have already been current, then to go on to present the « general statistical problem », and then to specialize to the problem of probability. We wish to make it clear at the outset, however, that an appreciable number of points appearing in this paper will be a repetition, in our own terms of course, of ideas that have already been published elsewhere. This repetition is, however, essential for the clarity of presentation of the material, because our principal conclusion is really that it is just the combination of these various ideas, which have appeared mainly in a scattered form until now, combined with our own results, which indicates

the possibility of a more general point of view toward statistical problems than has thus far generally been considered.

2. – Chance and Statistical Law.

The most important and characteristic application of the theory of probability is to a statistical aggregate of objects, events, or phenomena subject in some way to the action of « chance ». To know what probability means, it is therefore necessary to have a definition of chance. We shall give here such a definition (*).

In order to bring out what is meant by chance, we may consider a typical chance event; namely, an automobile accident. Now, it is evident that just where, when and how a particular accident occurs depends on an enormous number of more or less equally important factors, a slight change of any one of which could greatly change the character of the accident, or even avoid it altogether (+). Hence, relative to a context in which we consider, for example, the precautions that can be taken by a particular motorist, each accident has an aspect that is *fortuitous*. By this, we mean that what happens is *contingent* on independent factors occurring outside the context in question, which have no essential relationship to the characteristic traits that define just what sort of a person this motorist is and how he will behave in a given situation. For this reason, we say that relative to such a contest, a particular collision is not an inevitable development, but rather that it is an accident and comes about by chance from which it also follows that *within this context*, the question of just where, when, and how such a collision will take place, as well as that of whether it will take place or not, is unpredictable.

Let us now consider a series of similar accidents. Here, we note that there is a complicated and unpredictable variation or fluctuation in the precise details of the various accidents (e.g., precisely when and where they take place, precisely what is destroyed, etc.). The origin of this variation is easily understood, since as we have seen, the numerous causes on which the accident depends are essentially independent of each other and of anything that appears in the context of the actions taken by the particular motorist under discussion. Since

(*) After we had finished this work, Prof. MARIO SCHÖNBERG called our attention to the fact that some of the ideas on chance and probability appearing in this paper had already been suggested by COURNOT (*Exposition de la théorie des chances et des probabilités*, Paris, 1843).

(+) E.g.: one of the motorists could have started ten seconds earlier or later, or slowed down to avoid a cat that happened to cross the road, or he could have given the wheel a slightly different turn, etc..

each of these essentially independent sets of causes will in general act at different and not systematically related times and places, and in different and not systematically related ways in all other respects, the net result will be just the complicated fluctuations in the details of different accidents which we have described above, fluctuations which will be unpredictable in the context under discussion (although they are in principle predictable in a sufficiently broad context).

As the number of accidents under consideration becomes larger and larger, however, new properties begin to appear; for one finds that individual variations tend to cancel out and regular statistical trends begin to show themselves. Thus, the total number of accidents in a particular region generally does not change much from year to year; and the changes that do take place often show a regular systematic trend. Moreover, this trend can be altered in a systematic way by the alteration of specific factors on which accidents depend. Thus, when laws punishing careless driving are enforced, or when regular inspection of mechanical parts is required, the mean accident rate in a given region always undergoes a downward trend. In the case of an individual motorist taking a particular trip, no very definite predictions can in general be made concerning the effects of such measures, since there are still an enormous number of causes of accidents that have not yet been eliminated; yet statistically, as we have seen, variations in a particular cause produce a regular and predictable trend in the effect.

The behavior described above is found in a very wide range of fields including social, economic, medical and scientific statistics, and a whole host of other applications. In all these fields, the following are the most important characteristics:

1) Each event or result depends contingently on a very large number of essentially independent causal factors lying outside the context under discussion.

2) These causal factors all have comparable orders of importance (*) (although they will not in general be exactly equal in their effects).

3) These causal factors vary sufficiently so that in a series of similar events, the events of interest fluctuate in a way the details of which cannot be predicted from any data that can be taken solely in the narrower context

(*) If the effects of a few of the factors are many orders of magnitude more important than the effects of all the others, combined, then *these* factors would be regarded as operating systematically. The event in question would still be contingent on these causal factors, but it would be inappropriate to describe their effects as being those of chance.

under discussion. It can be predicted, however, that in a long enough series of events, every possibility will be brought about in these fluctuations. Moreover, the more sensitively dependent the event of interest is on the external causal factors the more favorable are the conditions for the creation of chance fluctuations of appreciable importance.

4) In an aggregate containing many members, the chance fluctuations tend to cancel out, so that the average or long-run properties can be predicted approximately on the basis of *statistical laws*. Such laws can be applied without the need to go to the broader context in which one would take into account the additional causal factors that govern the details of the fluctuations of the individual members of the aggregate.

3. – The Concept of Probability

Historically, the concept of probability was first given a precise form in connection with gambling games. A good example is furnished by the game of dice. Now, if we follow the results of each individual throws of the dice, we discover that the results fluctuate from one throw to the next, in the same characteristic way that is demonstrated by a wide range of chance phenomena of the type discussed in Sec. 1. Nevertheless, gamblers have developed the custom of betting on a given combination and of giving certain odds that depend on the combination in question. Experience has shown that corresponding to each possible combination, there seem to exist a set of « fair odds », such that if these odds are offered, then in the long run the gambler will neither win nor lose systematically.

Now the problems that was attacked by the earliest mathematicians (*) who turned their attention to this question was to find a theoretical way of calculating what these « fair odds » should be. In the case of throws of a die, for example, this problem was solved by supposing that all six faces of each die are « equally likely » in each throw. Thus, the probability that a given die will come out a five is 1/6, and since the dice are « independent », the probability that both will come out fives is $1/6 \times 1/6 = 1/36$. Hence, the « fair odds » are 36 to 1 in this case.

Although the method of solution of the problem indicated above certainly worked in connection with games of chance, it involved the introduction of the rather vague notion of equal « likelihood » or « equiprobability » of the various possible results of a throw. This notion contains a mixture of two

(*) The earliest mathematical discussions of probability seem to have been given by Pascal, Fermat, Huyghens, Bernoulli and Laplace.

very different interpretations of probability, which we may call respectively the « subjective » and the « objective ». For the sake of clarity of presentation, we shall give here, however, the pure forms of these interpretations, which developed only some time after the concept of probability first began to be used systematically.

In the subjective interpretation of probability, it is supposed that probabilities represent, in some sense, incomplete knowledge or information concerning the events, objects, or conditions under discussion. Thus, in the game of dice, we have no way of knowing with certainty before the dice are thrown what the results of each individual throw will be (since these results are determined by the initial positions and velocities of the various parts of each die which are not accessible to us in practice). Hence, if the dice are symmetrically constructed, we have no knowledge favoring the suggestion that we will obtain any one side instead of another, and we therefore assign equal probabilities to each side. In this point of view, then, probability is regarded as something that measures or reflects our degree of knowledge or ignorance. It is therefore essentially a description of a state of mind, and as a result, an essentially subjective category, since it would cease to be necessary, or even to have any meaning, if we could obtain precise knowledge concerning the initial motions of the dice in each throw.

It is clear, however, that the interpretation of probability as nothing more than a description of our mental reflexes under conditions of incomplete knowledge is not adequate in typical statistical problems, such as those arising in applications in science and in other fields. For such an interpretation gives us no notion at all of why in statistical aggregates containing many members, the actual *relative frequency* is generally close to numerical value of the probability (*). The mere fact that we do not know any reasons concerning either the structure of the die or its initial motion that would favor one face of a die over another evidently does not by itself necessarily imply approximately equal relative frequencies for each face. Thus, among the things that we do not know about the initial motions of the die, there could conceivable exist hidden factors tending to favor one result over another, even if we know that the die is well-balanced in its construction. Such tendencies could evidently exist whether we knew of them or not; and indeed it is just the nonexistence of these tendencies that is necessary for the characteristic regular

(*) The importance of the relative frequency in connection with probability was first recognized by Ellis (*On the Foundations of the Theory of Probabilities*, in *Trans. of Camb. Phil. Soc.*, 8, 1-6 (1844)), (paper read in 1842), and then by Cournot (*Exposition de la théorie des chances et des probabilités*, Paris, 1843). However, neither of these authors fully eliminated the subjective point of view. Venn (*The Logic of Chance*, London and New York, 1866) advocated systematically a pure frequency interpretation.

behavior of the relative frequency in the long rung. Thus, we have to do here with an *objective* property of the die and of the process by which it is thrown (*), which has no essential relationship to the degree of knowledge or ignorance of this process that may exist in our minds. Indeed, even if we somehow were able to obtain perfect knowledge of the initial conditions in each case, this would not change the fact that in a typical long series of throws, the probabilities would still provide a good approximation to the relative frequencies predicted on the basis of such perfect knowledge, (provided of course that the die is well-balanced and thrown by a process that leaves room for a sufficient degree of chance fluctuation in the initial conditions). Thus, the probabilities are objective properties of the die and of the process by which it is thrown, which manifests itself approximately as a relative frequency in a long series of throws. However, the relative frequency does not approach uniformly, but fluctuates within bounds that tend to come closer and closer to each other as the number of terms increases.

It is clear, that the objective interpretation of probability described above is the one that is appropriate in statistical problems such as those treated in science and in other fields. Of course, the word « probability », as commonly used, can also be given a subjective meaning in terms of our judgements regarding the likelihood that inferences based on incomplete information are valid. Nevertheless, as we have seen, the predictable regular trends in the relative frequencies are present in statistical aggregates of objects or events distributed by chance processes, whether we make inferences about them or not; so that the introduction of our degree of knowledge or ignorance is quite irrelevant to the problem under discussion.

With regard to the role of chance in providing the general conditions or background in which the concept of probability can correctly be applied, this follows immediately from the discussion of the origin of statistical law given in Sec. 2. Thus, in the case of the throws of a die, we have all of the characteristic factors tending to favor the creation of chance fluctuations. First of all, there is the relatively great sensitivity of the die to the precise character of the initial motions with which it is thrown. If the die is thrown from an appreciable height, so that it has time to turn one or more times before it falls, then a small change of its initial conditions could change the final result completely from any one face to any other. Moreover, the human body is evidently a very complex system, containing a great many fluctuating elements of comparable orders of magnitude. Thus, it is quite natural that these fluctuations are such that in the long run, any face of the die can appear.

(*) Both COURNOT and VENN recognized this fact explicitly. Indeed, the exposition in this section follows closely along the general lines laid down by COURNOT.

And considering the effective *independence* of the behavior of these elements relative to the physical state of the die, it is hardly surprising that in the long run and in the average, these fluctuations favor no particular face, and lead to approximately equal relative frequencies for each. We see then in detail how the property of probability is a reflection of various aspects of the entire process by which a distribution of chance events comes into being. Thus, the probability is a property of this process as a whole, which can be given a meaning even prior to the realization of any particular sequence as the property of the process as a whole from which the predictable trends in the relative frequencies come.

Finally, a word of caution must be given with regard to the use of the theory of probability. For it must be remembered that the theory is valid only to the extent that the distribution of objects or events has characteristics similar to those that would be brought about by the operation of a set of genuine chance fluctuations. Such fluctuations depend, for their characteristic properties, however, on the *independence* of the external and contingent causes relative to the events under discussion. Now in reality, because everything in the universe is interconnected with everything else to some degree, perfect independence cannot exist. Nevertheless, in many cases, such as that of throws of a die, the degree of interconnection is so slight, that the concept of *pure chance* provides a good approximation. Where the independence is incomplete, however, this concept will apply only to a limited degree of approximation; and then, many features of the calculus of probability (e.g. the multiplication of the probabilities of two events to give the probability that they shall both occur) will also apply only approximately. In such cases, we do not have pure chance, so that a more precise representation of the statistical laws would in general have to go outside the framework of the theory of probability (*).

4. – The Irregular Collectives of von Mises.

The objective interpretation of the theory of probability was given a new and more precisely defined mathematical formulation by VON MISES (1). Since this interpretation raised the interesting problem of randomness, and moreover had an important influence on modern thinking on the subject, in other respects, it will be worth our while to discuss and analyze this point of view briefly.

The basic idea of VON MISES was to introduce the notion of the *irregular*

(*) We shall discuss this problem in more detail in Section 9.

(1) See, for example, R. VON MISES: *Wahrscheinlichkeit, Statistik und Wahrheit*, dritte Aufl. (Wien, 1951).

collective (or statistical ensemble) as a model for a long series of chance events, or for a distribution containing many objects with properties that are subject to chance variations. This collective was given the following three essential characteristics:

1) It was assumed to have an unlimited number of members.

2) In the collective, the relative frequency of every type of object or event was assumed to approach a definite number, which was *identified* with the probability of that type of object or event (i.e., the *only* meaning of probability was assumed to be in terms of this relative frequency).

3) The distribution of objects or events was assumed to have a certain property of « randomness » or « lawlessness ».

To understand the suggestions of VON MISES with regard to the concept of randomness, it is convenient to consider the problem of *statistical sampling*. Now, it is well known that in typical distributions of chance events, one can draw partial samples by various methods, and that the relative frequencies of various classes of events in the samples provides a good approximation to those in the sequence as a whole. The suggestion of VON MISES is effectively then that the possibility of drawing samples of this kind be used as a definition of randomness. More precisely, his suggestions amounted to the idea that in a random distribution, subsequences drawn out of the collective in suitable ways should have the same limiting relative frequencies for each type of event as are present in the main sequence (*). Evidently, in a very regular sequence, this would not be so. Thus, if instead of throwing a die, we *placed* it on the table in the sequence 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6,..., etc., the limiting relative frequency for each possible result would clearly be 1/6. Nevertheless, if one sampled every sixth throw, one would always get the same result, so that the relative frequency of that result would be unity, and not one-sixth, as it is in the sequence as a whole. Thus, regular sequences of this type fail to possess certain important properties with regard to sampling that experience

(*) Actually, VON MISES developed his definition originally in connection with the problem of winning a gambling game by betting on subsequences of plays chosen from the main sequence by some predetermined kind of rule or « system ». The fact that in real games, such efforts to « beat the game » have always failed is then taken to demonstrate the randomness of the sequence of plays. It is evident, however, that the possibility of finding a « winning system » in a gambling game depends on whether or not samples drawn from the entire sequence of plays will have the same distribution of relative frequencies as are present in the sequence as a whole. Thus, the formulation of the definition of randomness in terms of the possibility of drawing representative samples contains the gambling game as a special case.

has shown to be essential for a random series of events; and it is for this reason that VON MISES proposes the identification of the property of randomness with irregularity, or « lawlessness ».

The infinite collective of VON MISES is essentially a simplified model for typical finite but long sequences of chance events. In other words, it is supposed that if one replaces the finite collective by the infinite model, then conclusions drawn by using the model will generally apply approximately to the real finite collective. The infinite collective is evidently an abstraction, which cannot exist in reality. But it is assumed by von MISES to be an abstraction which enables him to represent in a simplified or « economical » form the *essential* relationships that exist in a sufficiently large number of chance events. In this regard he compares his idealized abstract collectives to the introduction of the concept of points in geometry, which he likewise regarded as abstractions permitting a simplified or economical representation of the essential aspects of the geometrical relationships of real objects that are very small.

The formulation of probability given by VON MISES can be criticized on several grounds. First of all, one can raise the question as to whether it is adequate to identify probability *completely* as VON MISES does, with a limiting relative fequency in the idealized collective. For from the fact that the probability makes possible approximate predictions of future relative frequencies, one would be led to conclude that there should exist some property corresponding to the probability, even *before* the real collective has come into being. For example, let us consider a game of cards in which the relative frequency with which a given card will be obtained in a long series of plays is determined approximately in terms of a suitable probability. The question then immediately arises « From what properties existing before the cards are drawn does the approximate predictability of the relative frequency come? ». Surely, it cannot come from the fact that the plays of the cards that are potentially capable of being made belong to an idealized but non-existent collective, which at best may be realized approximately only in the future. Hence, the probabilities ought not to be identified with *nothing else but*, relative frequencies in the idealized ensemble, if only because such an identification implies that the approximate predictability of relative frequencies in real ensembles that are going to be produced in the future comes out of *nothing at all* existing in the present. Such an implication is evidently unsatisfactory, since it does not take into account the fact that definite conditions in the present are actually needed to lead to predictable relative frequencies in the future. Thus, in the case of the card plays, the probabilities will give the right answers only if the cards are similar in structure and if they are adequately shuffled by some chance process. Hence, it is evident that the approximate predictability of the relative frequencies comes out of something else existing at present. One is therefore led to conclude that the concept of probability needs a broader

base than can be obtained by identifying it with nothing else but a relative frequency in an idealized ensemble.

In this regard, the point of view of VON MISES is a step backwards relative to that of Cournot, who had already made it clear that even before an aggregate of objects or events is realized, the probability reflects various aspects of the objects and the processes from which the distribution will eventually come.

The second important criticism of the von Mises theory of probability is connected with the proposed criterion for randomness. The essential problem here is to define more precisely what is meant by « irregularity » or « lawlessness ». Thus, VON MISES already realized that he could not require that the relative frequency in *every possible* subsequence be the same as in the main sequence, since one could, for example in the case of throws of a coin, always choose that subsequence in which the results came out heads, so that the relative frequency of heads would be unity, and not one-half, as it is for the sequence as a whole. He dealt with this problem by requiring, that the subsequence be chosen by a method of sampling that is defined *before* all the results of a given sequence are available: or more precisely, that the subsequence be chosen by a method that depends only on the position of a given event in the sequence and on the properties of the events before it. Even so, however, the criterion is very vague, and cannot be applied in any practical problem (e.g. one could hardly count the relative frequencies in all possible subsequences fitting into the above criterion). Moreover, it is by no means clear that such a criterion is self-consistent, since it hardly seems likely that *every* method of sampling falling within its scope will give the same relative frequencies as in the main sequence.

The problem of defining what one means by « irregularity » or « lawlessness » is a particularly thorny one, since in the very act of defining it, one is inevitably ascribing to it some kind of regularity and law. Indeed, various members (*) of the school of VON MISES seem to have given up the attempt to define randomness as lawlessness. Taking as a starting point a sequence of plays in a gambling game, they propose to define randomness in terms of a specific and well-defined property of such a sequence: viz, that if a player systematically follows any one of a countable infinity of rules of play which they define, he can never « beat » the game. These efforts to define randomness are in their turn, however, subject to important criticisms (2). Aside from sharing in the defect of the approach of VON MISES that it is not clear how they could be applied in practice to a wide range of problems, those criteria

(*) A. WALD, A. H. COPELAND, etc..

(2) J. VILLE: *Étude Critique de la Notion de Collectif* (Paris, 1939).

for randomness involve a simple transference to all possible statistical problems of a method that may perhaps be appropriate in the special problem of gambling games (*). Finally, there has not yet been given a demonstration that these criteria can consistently be applied in all of the domains where the theory of probability has been used.

5. – The Modern Axiomatic School.

The modern axiomatic school of probability developed by KOLMOGOROFF (3) and others is based on essentially the same general assumptions concerning chance and statistical law that we have discussed in Secs. 2 and 3. Its essential new characteristic is its stress on taking the laws of the theory of probability as mathematical abstractions, which can be studied separately from the question of the problem of their interpretation, and of their applicability to real problems.

In this approach, the probabilities are assumed without further analysis as «primitive» mathematical concepts, while the relationships between them are then taken as axioms. The purpose of the mathematical theory of probability is then regarded as being the deduction of the logical consequences of these axioms in progressively greater detail. This procedure is compared with that used in geometry, where one likewise introduces certain abstract entities (e.g., points and lines) as primitive concepts not subject to further analysis, and where the function of the theory is likewise conceived of as the deduction in ever greater detail of the logical consequences following from a certain assumed basic set of axioms, by which it is assumed that the primitive concepts are related.

The principal advantages of the axiomatic point of view is then that without identifying probability prematurely and excessively narrowly with any specific property, such as relative frequency either in a real or in an idealized ensemble, it enables one to obtain a clear and precise formulation of the purely mathematical part of the theory.

Certainly if one simply considers such a process of abstraction in itself, no criticism whatever could possibly be made against carrying it out within the domain of pure mathematics, provided that it is clearly understood that

(*) As we shall see in Section 9, there are strong reasons to suppose that the criteria for randomness may have to possess a much wider range of adjustability to the specific characteristics of the various statistical problems to which they are applied, than is possessed by the criteria that we are discussing here.

(3) A. KOLMOGOROFF: *Grundbegriffe der Wahrscheinlichkeitsrechnung*, in *Erg. d. Math.*, **2**, n. 3.

perhaps under other conditions and in connection with new kinds of statistical problems, new kinds of mathematical abstractions may be needed (Just as the abstractions of Euclidean geometry may have to be replaced by the more general ones of non-Euclidean geometry, to deal with astronomical domains of space). However the position taken by many members of the modern axiomatic school seems to imply that the calculus of probability already represents the most general possible purely mathematical means of treating statistical problems. Such a point of view tends to confine the horizon of our thinking about statistical problems to nothing more than the calculation of one kind of probability in terms of another kind. On the other hand, as we shall see in the later sections of this paper there exists a wide variety of actual cases in which arise much more general statistical problems that can advantageously be treated by methods going beyond those of the theory of probability and that in many cases must be treated by such methods. Thus, it is our opinion that while the current theory of probability may consistently be regarded as a self-contained logical discipline, it may also be studied both mathematically and in more general ways, within the context of a broader point of view, in which the meaning of its basic concepts becomes clearer, as one comes to see their setting in the general domain of statistical problems, as well as the processes by which the laws of probability come to be valid, and the factors limiting the domains of validity of these laws.

6. On the General Statistical Problem as It Arises in Statistical Mechanics.

Thus far, we have discussed the treatment of statistical problems in terms of the concept of probability. We shall now see that in the domain of statistical mechanics, a much more general notion of what constitutes a statistical problem is not only advantageous, but in fact, in many problems, absolutely necessary.

In order to help bring out our point, we first review a few elementary ideas of statistical mechanics. Now, as is well known, the basic object of statistical mechanics is to predict various macroscopic properties of a system composed of a very large number of molecules (of the order of 10^{23}), on the basis of suitable hypotheses concerning the motions of its constituent molecules. To describe the macroscopic properties, we may divide the six dimensional space having $(\boldsymbol{x}, \boldsymbol{p})$ as coordinates (which was called μ space by EHRENFEST) [4] into cells that are so small that $\boldsymbol{x}$ and $\boldsymbol{p}$ do not change appreciably within them. Now if we consider a system composed of many molecules, the *macroscopic*

---

[4] P. and T. EHRENFEST: *Encyklop. der Math. Wiss.*, **4**, 4 (Leipzig, 1907-1914).

properties of this system can be specified by giving the number of molecules N_r, in each cell. Such a specification involves giving a series of numbers $(n_1, n_2, \ldots n_i, \ldots, n_s) = [n_r]$. In typical problems, the total number of molecules in the system is so great that the $[n_r]$ are all fairly large, even when the cells are chosen small relative to macroscopic orders of magnitude.

On the other hand, the basic laws of mechanics work in terms of a *microscopic specification* of the state of the system. This is done by specifying a point in the $6n$ dimensional phase space of the system as a whole (called the Γ space by EHRENFEST).

An important question that we now consider is that of computing the volume of Γ space corresponding to a certain set of the $[n_r]$. As is well known, this problem is of crucial importance in calculating the statistical properties of the system (5). This volume is

$$\Omega[n_r] = \frac{n!}{n_1!\, n_2! \ldots n_s!}\, \Delta\Omega\,, \tag{6.1}$$

where $\Delta\Omega$ is the volume in phase space of the cell corresponding to any one arrangement leading to the same $[n_r]$.

We now consider, as is usually done in statistical mechanics, a region of phase space lying between two hypersurfaces of constant energy, E, and $E + \Delta E$, where ΔE is quite small compared with E. As is well known, an isolated system moves along a surface of constant energy, so that if we study the properties of such a shell of thickness ΔE and let ΔE approach zero, we shall obtain an accurate representation of the properties of the hypersurface on which such a system moves.

The next problem is to obtain an approximation to the expression (6.1), which is valid when the n_i are large. This is conveniently done in terms of the expression for $\ln \Omega[n_r]$. We use Stirling's approximation for the n_r, obtaining

$$\ln \Omega[n_r] \approx n \ln n + \ln \Delta\Omega - \sum_i n_i \ln n_i\,. \tag{6.2}$$

Now, $\Omega[n_r]$ has a very steep maximum. Thus, as we shall see presently a good approximation to this function can be obtained by evaluating the set of values of $[n_r]$ for which $\Omega[n_r]$ is a maximum, and then by expanding $\ln \Omega[n_r]$ as a series of powers of $\delta n_i = n_i - n_{i0}$, where n_{i0} is the value of n_i for which $\Omega[n_r]$ is a maximum.

(5) This problem has been treated in various ways by many authors. See, for example, P. and T. EHRENFEST: *Encyclopedie des Sciences Mathématiques*, 4, 2, Supplément II.

To obtain the maximum of $\ln \Omega[n_r]$, subject to the condition following from the isolation of the system, that the total energy is a constant, $(\sum_i n_i E_i - E = 0)$ and that the total number of particles is a constant $(\sum_i n_i - n = 0)$ we use the well known method of Lagrange's parameters, i.e.,

$$\ln \Omega' = \ln \Omega - \lambda(\sum_i n_i - n) - \mu(\sum_i n_i E_i - E). \tag{6.3}$$

The condition for a maximum is then

$$\frac{\partial}{\partial n_i} \ln \Omega' \approx -\lambda - \mu E_i - \ln E_i = 0\,, \tag{6.4}$$

or

$$n_i = n_{i0} = A \exp[-\mu E_i]\,. \tag{6.5}$$

As is well known from calculations carried out in statistical mechanics μ comes out equal to $1/kT$, where k is Boltzmann's constant, and T the temperature, while A is a normalizing constant chosen to yield $\sum_i n_i - n = 0$.

For the case of molecules with negligible interactions, to which we shall restrict the discussion, the energy is just the kinetic energy, $E_i = \frac{1}{2} m \boldsymbol{v}_i^2$ so that we obtain

$$n_i = A \exp[-m\boldsymbol{v}_i^2/2kT]\,. \tag{6.6}$$

This is the well-known Maxwell distribution of velocities.

We see then that the Maxwell distribution corresponds to the largest possible volume of Γ space. We shall now be interested, however, in the volume of Γ space corresponding to small deviations from the Maxwell distribution. To calculate these volumes, we expand eq. (6.3) up to second powers in $\delta n_i = n_i - n_{i0}$ obtaining

$$\ln \Omega' = \ln \Omega_0' - \tfrac{1}{2} \sum_i (\delta n_i)^2 / n_{i0}\,, \tag{6.7}$$

$$\Omega' = \Omega_0' \exp[-\tfrac{1}{2} \sum_i n_{i0} (\delta f_i)^2]\,, \tag{6.8}$$

where $\delta f_i = \delta n_i / n_{i0}$ is the fractional deviation from a Maxwell distribution.

We see then that deviations from a Maxwellian distribution have a gaussian density in phase space. Indeed, because n_{i0} is large, the volume corresponding to appreciable fractional deviations is, in general, very small. In other words,

the maximum of $\Omega[n_r]$ is a very steep one (*). Indeed, it is so steep that one may express the result more pictorially by saying that a surface of constant energy corresponds to a Maxwellian ocean in which there are a few very tiny non-Maxwellian islands (+). Now, the usual treatment in statistical mechanics has been to assume that every point on the shell of phase space, in which the energy has a value between E and $E+\Delta E$ is equally probable. Such an assumption would evidently lead to the correct conclusion that the system practically always has a distribution very close to the Maxwellian: for the very small volume of phase space corresponding to the non-Maxwelllian distribution would then practically never be entered by the phase point.

The question of the justification of the hypothesis of equiprobability on a shell of constant energy is discussed in texts on statistical mechanics (6). This question is, however not primarily what interests us in this paper. Hence, we shall here content ourselves with the statement that such a justification usually involves attempts to prove either a quasi-ergodic theorem or metric-indecomposability of the orbits in phase space, attempts which thus far have not been successful except for a few of the simplest types of systems (−). The point that interests us here, however, is that the attempt to prove equi-probability may have little or no relevance to the problem of demonstrating the Maxwellian distribution. For because of the fact that the non-Maxwellian distributions correspond to such a minute volume of phase space, it is clear that almost any continuous probability distribution, uniform or otherwise, would lead to a distribution very close to a Maxwellian distribution. But more important even than this is the fact that a very wide range of *determinate laws governing an individual system* would still give an approximately Maxwellian distribution, for practically all of the time, as long as they did not imply an unusual tendency for the orbit to remain near the special non-Maxwellian regions. And since as we shall see in Sec. **7**, we have a very large number of particles in unstable motion, it is quite plausible to expect that in systems of the kind with which we work in statistical mechanics, the actual orbit in phase space will tend to wander sufficiently throughout the accessible regions of phase space, so that it will not long remain in the non-Maxwellian regions.

(*) A much more elegant and precise method is available for deducing this result. See, for example, the proof of the central limit theorem in A. I. KHINCHIN: *Statistical Mechanics* (New York, 1949). The method that we have chosen here is, however, based on essentially the same ideas, and has the advantage of bringing out more clearly what is important for our purposes.

(+) This fact has long been well known. See, for example, reference (5).

(6) See for example, A. I. KHINCHIN: *Statistical Mechanics* (New York, 1949).

(−) We shall discuss a few such cases in Section **9**.

We conclude then that the effort to base statistical mechanics on equiprobability in a given region in phase space is an unnecessarily narrow way of attacking the problem. In some cases (if there is time for the orbit to fill such a region to a good degree of approximation), it could be based on equiprobability (*). But in most problems, the real basis seems to be the insensitivity of large-scale averages in systems containing many components to the precise details of the motions of these components. Because of this insensitivity, one may have practically constant and determinate statistical averages, without having equidistribution, and indeed without having any kind of probability distribution at all (+). Thus, the problem of finding an adequate basis for statistical mechanics is greatly simplified, because it is evidently much easier to prove that the phase point does not remain long in the tiny non-Maxwellian regions than it is to prove that it covers the whole space in a quasi-ergodic or metrically indecomposable way.

On the basis of the above results, we are then led to consider a new and broader point of view on statistical problems, which permits us to go beyond the mere calculation of one kind of probability in terms of another kind. Instead, we may consider what has been called in Sec. 1 the «general statistical problem». This is to find which properties of a statistical aggregate are approximately independent of the variations that actually occur in the detailed behavior of the individual systems composing the aggregate under the conditions that define the problem. The proof of the quasi-ergodic theorem and theorems in metric indecomposability would then be special cases of solutions of this problem, because where these theorems are valid, the long-run properties of the orbit in phase space are essentially independent of initial conditions. Thus, predictions of the mean behavior become possible without a detailed specification of the state of an individual system.

We shall now further explain some of the consequences of the broader and more general attack on statistical problems that we have described above. In this connection, we have thus far considered two extremes, one of which is to deduce the statistical properties from probabilities, and the other of which is to deduce them from precise determinate laws governing the individual

(*) In the problem that we are considering here, there is not enough time for this to happen. Thus, as a consequence of a theorem proved by POINCARÉ (*Acta Math.*, **13**, 67 (1890)), in a typical mechanical system, containing something like 10^{22} molecules, it would take a fantastically long time, of the order of trillions of trillions of years for the orbit to come reasonably near to every point of phase space.

(+) This point seems to have been for a long time known to a number of writers in the field (see, for example, A. I. KHINCHIN: *Mathematical Foundations of Statistical Mechanics*, sec. 13). However, as far as we can tell its full significance with regard to the lack of necessity of using a probabilistic treatment of statistical mechanics does not seem to have been realized.

systems. We may, however, consider the possibility of deducing the statistical properties from intermediate types of laws, which are neither completely probabilistic nor completely deterministic. The need for such intermediate types of laws can be seen by considering the fact that no system can be described completely in terms of any finite number of properties. First of all, any system is always interacting with other systems outside of it. In many cases, these interactions seem to be small. But because the detailed motions of individual systems are unstable, they can in the long run, be affected a great deal even by small disturbances arising outside. Secondly, laws operating at a particular level are generally subject to disturbances arising at other levels, not included in the laws in question. For example, the laws of the macroscopic level are subject to fluctuations of the type of Brownian motion, arising from the motions at the atomic level. But the laws of classical mechanics operating at the atomic level are subject to further fluctuations arising from the quantum-mechanical properties of matter. And because of the instability of motion, these fluctuations can have important effects on the detailed behavior of the orbits even at a level where classical mechanics would otherwise be a good approximation. Thus, one comes to the conclusion that in any real problem, in which one necessarily leaves out of consideration an unlimited number of effects coming from outside from the system and from other levels, a precisely determinate and complete causal prediction cannot be obtained. In general, therefore, we expect to obtain from our laws predictions only of the mean behavior of a system, and of the range and general character of the fluctuations away from this mean, arising from the infinity of factors that have not been taken into account. Because of the insensitivity of large scale averages in statistical mechanics, however, a very wide range of laws of the type described above is evidently possible, which lead to approximately the same large scale statistical properties (such as the Maxwellian distribution).

On the basis of considerations of the type discussed above, one is led to consider the question of whether one ought not to seek a more general statistical formalism (*) and not always to try to frame all statistical problems in terms of the calculus of probabilities. Of course, the theory of probability has demonstrated its applicability in a wide range of fields, starting with gambling games, and going on to the fields of medical, social, industrial, economic, and scientific statistics. Nevertheless, as we have already seen, in our discussion of the foundations of statistical mechanics (certainly a study that falls within the scope of statistical problems), the theory of probability already

(*) This problem would be a mathematical reflection of the general statistical problem that we are raising here. The notion that new statistical formalisms, other than the currently used calculus of probabilities, might be possible was first suggested to us by Prof. Mario Schönberg (Private communication.).

seems to have little relevance. To be sure, it is possible to justify statistical mechanics on the basis of the hypothesis of equiprobability in phase space, but in most problems, this hypothesis can be shown not to have much to do with the real problem (*). What is significant in the real problem is that large scale averages are insensitive to most of the details of the molecular motions, so that for a wide range of possible motions, approximately the same macroscopic averages are obtained. The fact that one may within a certain context of problems, obtain correct results using the irrelevant hypothesis of equiprobability, is then a consequence of the close relationship between the equidistribution of probability that would arise in an extremely long period of time and the properties of the motion that lead to the statistical laws that are valid over periods of time of practical interest. One may compare the probabilities that would apply in an extremely long period of time to a sort of shadow, the behavior of which can more or less faithfully represent some of the statistical properties that are relevant in shorter periods of time. Our point of view is then that it could be profitable to study the essentials of the problem more directly rather than to study nothing more than the probabilistic « shadow ».

As an example of how statistical problems can be treated in ways other than those based on the theory of probability, let us consider the problem of finding the meaning of the entropy on the basis of the molecular properties of the system. Now, it has been customary to define entropy in terms of probability; and for a broad range of purposes, this definition has served to make possible the deduction of the thermodynamic relations and the calculation of some of the thermodynamic properties of various systems. But from the work of P. HERTZ and LEVI-CIVITA [7] one sees that entropy may be associated with mechanical properties of the system rather than just purely probability. Thus it seems quite likely that new models for the entropy could be worked out. We do not wish to imply that either the probabilistic or the mechanical model necessarily exhausts all that there may be to the concept of entropy. Indeed, still other models may perhaps be proposed later which give us even deeper insight into the meaning of entropy. But the mere fact that models different from the probabilistic one can lead to similar results in a broad domain is a sign that it would not be wise to identify entropy completely with the current probabilistic model, just because this model has led to correct results

(*) Thus, because of the extremely great length of the Poincaré cycles, there is no time for a system to come reasonably close to every point on a surface of constant energy in phase space.

[7] P. HERTZ: *Weber-Gans*, in *Repertorium d. Physik*, Bd. 1, 2 (Leipzig, 1916); T. LEVI-CIVITA: *Drei Vorlesungen über adiabatische Invarianten*, in *Abh. d. Math. Seminars Hamburg*, 6, 323 (1928).

in a certain domain. Rather, the way should be left open for new statistical concepts, and associated mathematical methods, which may shed new light on the meaning of entropy.

Another important application in which the concept of probability may have only a rather limited validity is the study of non-equilibrium problems in statistical mechanics. If we are interested in how a system approaches statistical equilibrium, then a certain difficulty of principle arises in connection with the application of the concept of probability. For the statistical properties of a system, (e.g., pressure, temperature, velocity, etc.), are changing with the time. The usual method of defining probability as the average over an unlimited period of time then will not work. Of course, if the statistical averages are changing slowly, one may define the probability approximately in terms of the average over a period of time during which these properties do not change very much (8). For rapid changes, however, even this method does not work. The other possible method of defining probability is through a statistical ensemble. But it seems unsatisfactory to have to treat what is after all the approach of a single system to equilibrium in terms of an average over a non-existent ensemble. Such a treatment may give correct results in certain contexts, but again it is rather like studying the motion of a thing in terms of the motion of its shadow. Perhaps new statistical methods and concepts would here make possible a deeper understanding of the problem of the approach to equilibrium in such systems.

Even if we restrict ourselves to statistical laws that are expressible in terms of the laws of probability, our point of view shows that one can still treat a broader range of problems than just the calculation of one kind of probability in terms of another. For within the setting in which statistical law is regarded as more general than probability, it has meaning to demonstrate the validity of the laws of probability in certain contexts, and to calculate the numerical values of the probabilities, on the basis of laws that are more general than the laws of probability. Thus, as we have seen, we can calculate a probability from a determinate law, or from an intermediate type of law, which is partially determinate. Vice-versa, a partially determinate law can be obtained from a probabilistic law, when the number of elements N, in the distribution is large. For example, according to Bernoulli's theorem, one can for practical purposes predict that the error in a certain average will decrease as $N^{-\frac{1}{2}}$. As N approaches infinity, the partially determinate law will become more and more nearly a completely determinate law. Thus, we see how the adoption of a broader point of view toward statistical law permits us to attack the problem of the interconnection of various kinds and levels of law.

(8) See, for example, M. SCHÖNBERG: *Nuovo Cimento*, **10**, 419 (1953) where a discussion is given of the problem of taking averages in a non equilibrium process.

7. – A Further Example of the General Statistical Problem - The Deduction of a Probabilistic Law from a Determinate Law.

In the previous section, we have seen in terms of an example drawn from statistical mechanics, in which one deals with a system containing a very large number of particles, how statistical relationships can be treated without the need for their being expressed in terms of the theory of probability. In this section, and in the next, we shall go to the opposite extreme and consider a system described by a single variable only, satisfying a simple determinate law, a law in which, however, the motion has the essential characteristic of extreme instability (*). We shall then see that this law leads to a long run behavior that is independent of initial conditions, a long run behavior that is just what is predicted by the theory of probability. We shall also see that as happened in the many-particle systems treated in statistical mechanics, essentially the same kind of long-run behavior will follow from a wide range of laws that are between the determinate and probabilistic types. Thus, the justification of the laws of probability in a given application is seen to be a part of the general statistical problem.

In order to motivate the example that we are going to choose, let us consider the path of a molecule in a gas, which is contained in a box. We suppose that this molecule can be represented approximately as a hard elastic sphere. Now, it is well known that in its collisions with other molecules, this particular molecule will follow a very complicated path, in wich the velocity undergoes a wide range of rapid fluctuations, both in magnitude and in direction. In the long run, however, the probability that this particular molecule will occupy any given small region of space is the same as that it occupy any other such region having the same volume. Likewise, the probability that the direction of motion lies in any specific range of angles becomes uniform in the long run.

Let us now analyze how such a uniform probability comes about. We first note that the path of any given molecule possesses an extreme instability. Thus, suppose the initial position of the i-th molecule to be $\boldsymbol{x}_i^0$, and its initial velocity $\boldsymbol{v}_i^0$. We now consider a collision of this molecule with any other one, say molecule j. Focussing our attention on the path of molecule i, we see that the change of angle of motion $\Delta\theta_i$, is evidently very sensitively dependent on the collision parameter, h, (shown in Fig. 1) and therefore on the initial position and velocity of this particle (+).

(*) POINCARÉ has already pointed out the importance of instability in this problem. We shall discuss POINCARÉ's work and compare his results with ours at the end of this section.

(+) It also depends on those of particle j, but in this discussion, we are interested only in following what happened to particle i.

In fact, if its initial position is changed by an amount of the order of the size of an atom, its angle of motion after collision will change by an amount of the order of 2π.

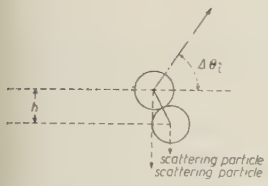


Fig. 1.

Since the change of position suffered by the particle between the times t_1 and t_2 is (as long as no collisions occur between t_1 and t_2)

$$\Delta \boldsymbol{x}_i = \boldsymbol{v}_i^0(t_2 - t_1),$$

it is clear that in a series of collisions, there is a continual multiplication of the instability, so that after a sufficiently large number of collisions, a change of initial position of less than the size of an atom can carry the particle to a very different region of space and to a very different direction of motion, as shown in Fig. 2.

Fig. 2.

It is already clear intuitively that under these conditions a very irregular motion will develop which does not favor any particular region of space. Thus the equiprobability of all regions of space becomes plausible on the basis of the kind of motion that exists.

It would evidently be desirable, however, to prove this equiprobability in a more rigorous way. Unfortunately, in the problem described above, the mathematical difficulties attendant on such an exact proof have not yet been resolved. We shall therefore propose here a simplified example, which permits an exact treatment, and which still retains the essential property of instability of motion, that is evidently at the root of the development of an equiprobability in the distribution of colliding molecules.

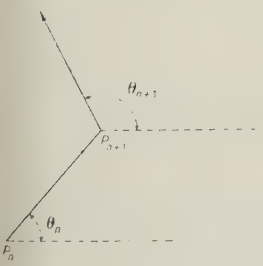


Fig. 3.

In this model, we consider a point, P_n, in a two dimensional space. The coordinates, x_n, y_n, represent the location of the n-th particle during the n-th collision, while θ_n represents its angle of motion after this collision, as shown in the figure below. We now assume that the particle, makes a new collision after moving a definite distance, L, (In reality, this distance L fluctuates from collision to collision; but by regarding this distance as a constant, we are further simplifying the problem.)

We further assume that in the process of collision, the angle θ_n is mul-

tiplied by a large integer, K, or that

$$\theta_{n+1} = K\theta_n . \tag{7.1}$$

We see that this hypothesis embodies the essential element of instability, since a small change of angle $\Delta\theta_n$ produces a much larger change of angle, $\Delta\theta_{n+1} = K\Delta\theta_n$. A change $\Delta\theta_n = 2\pi/K$ produces a change $\Delta\theta_{n+1} = 2\pi$. But, whenever the angle changes by 2π it comes back to its original value. Thus, we should write $\theta_{n+1} = K\theta_n - [K\theta_n]$ where $[K\theta_n]$ is the largest multiple of 2π in $K\theta_n$.

We see also that our simplified model gives a schematic representation of the significant aspects of the actual process of collision. First we note that the angle θ_{n+1} changes by 2π when the initial position of the particle just before it strikes the $(n-1)$-th obstacle changes by $L\Delta\theta = 2\pi L/K$. Now if K is large it is clear that a small change of position will lead to a change of angle of 2π. Thus we embody in our model something similar to the extreme sensitivity of the angle of collision to the collision parameter, that is so important in real collision processes for bringing about a uniform distribution. In fact, one could regard this model as more or less equivalent to having relative to each collision point a uniformly distributed set of obstacles on a circle of radius L, each of diameter $2\pi L/K$. Of course, in any given collision, the locations of the effective obstacles will depend on where the previous collision took place; whereas in the real process it does not. In this sense, our model is a further simplification; but in the present work, the essential aspect of instability of motion is the principal feature of the real process that we are interested in retaining in the model.

Let us now consider a sequence of collisions, starting with $\theta = \theta_0$. We have

$$\begin{cases} \theta_1 = K\theta_0 - [K\theta_0] , \\ \theta_2 = K^2\theta_0 - [K^2\theta_0] , \\ \cdots\cdots\cdots\cdots , \\ \theta_N = K^N\theta_0 - [K^N\theta_0] . \end{cases} \tag{7.2}$$

We then write our sequence associated with N collisions as

$$S(N, \theta_0) = \{\theta_0, \theta_1, \ldots, \theta_N\} .$$

Note that each θ_0 determines an *entire* sequence.

We now divide θ into elements of size $\Delta\theta$, of small, but fixed dimensions. And in each sequence $S(N, \theta_0)$ we can ask how many values of θ_n fall inside the i-th element (located near $\theta = \theta_i$). This quantity we define as $\varrho(N, \theta_0, \theta_i)\Delta\theta_i$.

Now, in order to demonstrate the development of a uniform distribution, what we must do is to show that for arbitrary θ_0 (with the possible exception of a set of measure zero) $\varrho(N, \theta_0, \theta_i)$ approaches a constant, independent of θ_0 and θ_i when N approaches infinity.

There exist well known theorems in the theory of numbers, which permit us to prove that in our case (as well as in a much wider class of similar problems). $\varrho(N, \theta_0, \theta_1)$ does indeed approach a constant (9). In the present paper, however, we shall demonstrate our results in another way which has the advantage of bringing out more clearly the essential features of the processes leading to equiprobability than is possible with the more abstract methods, such as those used by WEYL.

Let us first consider a sequence in which the particle arrives in the region $\Delta\theta_n$ after the n-th collision. We now ask: What would have been the values of θ before the n-th collision? One possible set of values could have been a region of width $\Delta\theta_{n-1} = \Delta\theta_n/K$ near $\theta_{n-1} = \theta_n/K$. But since the physically significant values of θ run between 0 and 2π the particle could also have come from any one of K regions of similar width located near $(\theta_n/K)+2\pi M/K$ where M is an integer running from 1 to $K-1$. Thus, the possible regions are uniformly distributed around the circle. (See Fig. 4). But each one of these regions $\Delta\theta_{n-1}$ could have come from a set of similar uniformly spaced regions of width $\Delta\theta_{n-2} = \Delta\theta_{n-1}/K = \Delta\theta_n/K^2$. If we continue this analysis back to θ_0, we discover that the values of θ_0 which can lead to a given $\Delta\theta_n$ are situated inside K^n uniformly distributed regions, each of width $\Delta\theta_0 = \Delta\theta_n/K^n$.

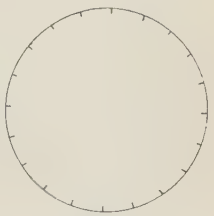


Fig. 4.

Let us now consider any *given* sequence, $S(N, \theta_0)$ where θ_0 lies in the range $\Delta\theta_0$, θ_1 in $\Delta\theta_1$, etc. For the sake of convenience, we choose all of our ranges $\Delta\theta_0, \ldots, \Delta\theta_N$ equal to each other and equal to $2\pi m/K$ where m is an integer which is much smaller than K, but large compared with unity. Thus, the unit circle is divided into K/m regions, each of width $2\pi m/K$.

We now ask « What is the range of θ_0 that can lead to any *given* sequence, in which the particle goes through the regions $\Delta\theta_1$, $\Delta\theta_2$, ..., $\Delta\theta_N$? ». This quantity, we call the « measure » of our sequence.

Now, as we have seen, there will be K uniformly distributed regions of width $\delta\theta_0 = \Delta\theta_1/K$ from which the particle started, if it arrived in the region, $\Delta\theta_1$. The total measure of these regions is $K\Delta\theta_1/K = \Delta\theta_1$. But now, if the particle goes through the region $\Delta\theta_2$, there are likewise K uniformly distributed regions of the angle θ each having width $\Delta\theta_2/K$, from

(9) See, for example, H. WEYL: *Math. Ann.*, **77**, 313 (1916).

which the particle could have started its second step. But of these, only m will be in the region of width $\Delta\theta_1$, in which the particle was actually located just before this second step. Hence, the measure of those trajectories passing through both $\Delta\theta_1$ and $\Delta\theta_2$ will be $m\Delta\theta_1/K = \Delta\theta_0 \cdot (\Delta\theta_0/2\pi)$. Similarly, the measure of those trajectories passing through $\Delta\theta_1$, $\Delta\theta_2$ and $\Delta\theta_3$, will be $\Delta\theta_0 \cdot (\Delta\theta_0/2\pi)^2$ and that of those passing through $\Delta\theta_1, \ldots, \Delta\theta_N$ will be $\Delta\theta_0 \cdot (\Delta\theta_0/2\pi)^N$. Thus, only a very small fraction of all possible initial values of θ_0 can lead to a sequence of angles passing through a given set, $\Delta\theta_1, \ldots, \Delta\theta_N$. The problem is very similar to that of having to get a ball to go through a series of irregularly spaced holes, if the ball is deflected as it passes through each hole by an amount that is very sensitive to the distance from the center. To get the ball through all of the holes, a fantastically good aim would be needed.

Our next step is to find the measure of those values of θ_0 that can lead to a given distribution, $N_i = \varrho(N, \theta_0, \theta_i)\Delta\theta_i$.

Now, as we have seen, any *given* sequence has a measure of $\Delta\theta \cdot (\Delta\theta/2\pi)^N$. We must now count all sequences that can lead to a given $\varrho(N, \theta_0, \theta_i)$ and therefore to a given set of N_i. This number is (*)

$$N!/N_1!\, N_2! \ldots N_i! \,.$$

Thus, the measure of those values of θ_0 leading to a given distribution of the N_i is

$$M = \frac{N!}{N_1!\, N_2! \ldots N_i!} (\Delta\theta/2\pi)^N \cdot \Delta\theta \,. \tag{7.3}$$

Using Stirling's approximation, we get

$$\ln (M/\Delta\theta) \approx N \ln N - \sum_i N_i \ln N_i + N \ln (\Delta\theta/2\pi)^N \,. \tag{7.4}$$

We obtain the result that the distribution having the highest measure is

$$N_i = \text{constant} = \bar{N} = N\Delta\theta/2\pi \,. \tag{7.5}$$

(*) Here, we make use of the fact that if any given sequence, $S(\theta_0, N)$ is possible then any other sequence involving an exchange of any two of the angles is also possible. This follows from the large value of the integer, K, which has as a consequence that given any values of θ_n, we can always obtain any desired value of θ_{n+1} simply by a small change of θ_n that does not take it out of the region $\Delta\theta_n$ of width $2\pi m/K$.

By defining $\delta N_i = N_i - \bar{N}$, we obtain

$$\ln M \approx \ln M_0 - \sum_i (\delta N_i)^2/2\pi = \ln M_0 - \sum_i 2\pi(\delta N_i)^2/2N\Delta\theta \,. \tag{7.6}$$

Writing for the fractional deviations, $\delta f_i = \delta N_i/\bar{N} = 2\pi\delta N_i/N\Delta\theta$ we get

$$M \approx M_0 \exp\left[-\sum_i \tfrac{1}{2}(\delta f_i)^2(N\Delta\theta/2\pi)\right] . \tag{7.7}$$

We see then that the measure of those distributions in which δf_i is appreciable becomes negligible as $N \to \infty$. Thus, fixing a given $\Delta\theta$ and a given δf_{i0}, however small, we can always choose N so large that the measure of those sequences in which $\delta f_i > \delta f_{i0}$ is less than any specified number.

What is the meaning of this result? We first remind ourselves that all of the possible sets of initial angles, that can lead to a given distribution, $N_i \quad \varrho(\theta_0, \theta_i, N)$ arrange themselves into small elements of width, $1\theta/K^N$. Thus as $N \to \infty$, the width of each element approaches zero very rapidly. This behavior corresponds to the extreme instability of motion, since a change of $\Delta\theta_0 = \Delta\theta/K^N$ would change θ_N by $\Delta\theta$ and therefore alter the distribution.

Thus, to get any *given* type of sequence, as $N \to \infty$ we would need a perfect fit for the boundary conditions. But as we have seen, the overwhelming majority of sequences correspond to distributions that are very nearly uniform. Only a comparatively small number of regions of width $1\theta/K^N$ will lead to anything different (such special regions will occur, for example, where θ_0 is rational). Although the number of such special regions approaches infinity as $N \to \infty$, this approach is much less rapid than is the decrease in width of each region, so that the total measure of such special regions approaches zero, as $N \to \infty$.

With regard to the theory of probability, the essential meaning of the above result is that the long run average distribution of angles is practically independent of the initial angle, θ_0. It is true that there is a special set of measure zero that could give rise to non-uniform long-run distribution, but one can show that such special values of θ_0 play no role in real problems. To see why this is so, we note that in order that one of these special non-uniform distributions should come about, it would be necessary that the value of θ_0 be defined *exactly* (i.e., with no error whatever). We know, however, that it is impossible in reality to have a physical system in which the value of a continuous variable is defined as *exactly* equal to any given number. For every system is always interacting with other systems in the environment which also undergo irregular fluctuations in the motions of their elements (such as atoms and molecules). Thus, even if a particle accidentally did have its initial value of θ_0 exactly equal to one of the special set of measure zero described above, it would be sure to obtain a slight disturbance from its environment; and because of the instability of motion, even the slightest disturbance would carry

it far from this special state and lead eventually to a uniform distribution. Now, it is true that such a disturbance might accidentally throw the system into another of the special angles, θ_0, of the set of measure zero, not leading to a uniform distribution. This would require however a very special exact coordination between the external disturbance and the motion of the system. It is evident that for practically all possible motions, this kind of coordination would not exist. Thus, it is consistent with the theory to suppose that the motions are such that there are no such coordinations. Moreover, this supposition is also consistent with all of our practical experience. For example, if we consider a cone standing on its point, this position is one of unstable equilibrium. It would be conceivable that the fluctuations of the surrounding air and of the table on which it is supported would be just such as to maintain it in its position of unstable equilibrium for all time but evidently this never happens in reality (*).

We conclude then that the treatment of the sequence given here is a special case of the general statistical problem, in the sense that certain long run or average properties can be predicted without taking into account the precise factors needed for a prediction of the details of the behavior of the system. For as we have seen, once we exclude the special set of θ_0 of measure zero described above, then the mean distribution of angles is independent of θ_0, and is therefore independent of the boundary conditions that determine the precise motion. We conclude then that the average value of any integrable function of θ, $\bar{f(\theta)} = \lim_{N\to\infty}(1/N)\sum_{n=1}^{N} f(\theta_n)$ will, independently of initial conditions, be equal to

$$\bar{f(\theta)} = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta\,. \tag{7.8}$$

The above is evidently equivalent to a statement of the equidistribution over angles. Thus, as far as computing averages of any function of θ are concerned, we obtain the same result as that which follows from using the theory of probability, with the probability distribution, $P(\theta)$ set equal to a constant.

We shall now compare our results with those of POINCARÉ (10), who treated a similar problem, but by different methods. POINCARÉ realized clearly the

(*) It is interesting to note that COURNOT in his *Exposition de la Théorie des Chances et des Probabilités* already realized the importance of this point, and in fact used the example of the cone standing on its point as an argument in favor of the thesis that in chance phenomena, no special properties depending on the infinitely precise specification of any particular quantity could be of physical importance.

(10) H. POINCARÉ: *Calcul des Probabilités* (Paris, 1912).

importance of instability in leading to an equidistribution. Indeed, he considered the case of a packet of asteroids, each in neighboring, but slightly different circular orbits. Because the phase angle of any particular asteroid relative to that of another is unstable (i.e., it increases indefinitely with the time), one expects that in time the asteroids will have a uniform and more or less random distribution of phase angles (which is in fact found in astronomical observations (*)).

In order to treat this problem, POINCARÉ considered an arbitrary *continuous* initial probability distribution (e.g., one having a very narrow but continuous peak near a certain point). He then showed that independently of the precise initial form of such a distribution, the probability would eventually become a constant that is independent of the angle. For discontinuous initial distributions (e.g., a delta function) this would not in general occur. But here POINCARÉ used an argument similar to the one by which we have excluded sets of measure zero as not representing real physical possibilities. He did this by pointing out that in all real problems, there is a background of fluctuations coming from influences outside the scope of what is treated in the problem (e.g., the motions of other planets, and even the molecular motions) so that no discontinuous probability function (such as a delta function) can represent any real physical situation.

POINCARÉ's conclusion was thus essentially the same as ours. But the weak point in his argument is that his mathematical methods do not correspond to the real context of his physical ideas. Indeed, from a strictly logical point of view, the assumption of the initial narrowly peaked and continuous initial distribution is not fundamentally different from making assumptions of the ultimate uniform probability distribution to begin with. It would thus be concluded that while POINCARÉ has demonstrated an interesting and useful relationship between two probability distributions taken at different times, he has still remained within the basic frame-work of just calculating one probability in terms of another.

On the other hand, the significance of our result is that the development of an equidistribution can be understood without any reference to the concept of probability at all. This we have done by recasting WEYL's results on the development of an equidistribution in a form in which the physical significance of the sets of measure zero could easily be seen, so that we can understand why they must be excluded (+).

(*) The rings of Saturn, made of small-meteor-like particles, present a similar problem.

(+) In section **10**, we shall see how an example similar to the one discussed by POINCARÉ can be treated, without the use of probability.

Finally, we shall consider the possibility of proving equi-distribution in terms of laws intermediate between determinate and probabilistic types, discussed in the previous section, in connection with statistical mechanical problems. It is important to consider such problems, because the notion of a single mechanical variable, following a determinate law, defined with unlimited precision, is evidently an abstraction. In reality, every such mechanical variable is subject to fluctuating perturbations of the type that we have mentioned in the previous section and in the above paragraph. In other words, it is in reality coupled, even if only slightly, to an infinity of other parameters. The extreme instability of motion could conceivably make such coupling significant over a long time. It is important therefore to show that there exists a wide range of cases in which the equidistribution would not be significantly affected by such perturbations.

In order to do this, we introduce the notion of the *sequence space*. The coordinates of a point in this space are specified by the values, $(\varphi_0, \varphi_1, \ldots, \varphi_N) = -[\varphi_n]$ of the various angles in any specified sequence that we wish to consider. Moreover, once we have chosen φ_0, then any point in the subspace $(\varphi_1 \ldots \varphi_N) = -[\varphi_n; \varphi_0]$, which starts at $\varphi = \varphi_0$ corresponds to a particular determinate relationship defining every term in the sequence in terms of φ_0. (The relationship $\varphi_n = K^n\varphi_0$ is evidently a special case). By considering a range, $\mathrm{d}\varphi_1 \ldots \mathrm{d}\varphi_N$, of points corresponding to a given φ_0, we are effectively considering a range of possible relationships to determine the sequence.

Now, as we did in our discussion of statistical mechanics in Sec. **5**, we can define the *range of sequence space* corresponding to a given distribution of the angles, θ (instead of the range of *phase space* considered in Sec. **5**). We need not go through the mathematical details here, as they are very similar to what was done in the statistical mechanical problem, but we shall merely quote the result here.

An equidistribution of the angles, θ, corresponds to the maximum possible volume of sequence space (*), distributions near the uniform one corresponding to a volume of sequence space that decreases according to a Gaussian curve. When N is large, this decrease is so rapid that almost all of sequence space corresponds to a uniform distribution.

The above result has a number of interesting consequences. First of all, we see that the special law $(\varphi_n = K^n\varphi_0)$ that we are discussing is not essential to an equidistribution, so that there are a wide range of determinate laws

(*) In the statistical mechanical problem, the Maxwellian distribution, instead of the uniform one, was obtained as the one corresponding to maximum volume in phase space, because of the requirements of conservation of energy. No analogous requirement appears in the present problem.

that could lead to the same statistical results. Indeed, even a law intermediate between determinate and probabilistic could have the same statistical effects, provided that as N increases without limit, it is quite rare that the point in sequence space finds itself inside the minute volume corresponding to distributions deviating appreciably from a uniform one.

The above result must, however, be interpreted with care. It does not by any means signify, for example, that all distributions must be uniform. Laws are evidently quite possible that would lead to non-uniform distributions, or even to distributions that approached no particular limiting behavior at all. All that we have shown is that a wide range of laws exist that could lead to a uniform distribution, so that the appearance of such a distribution need not be sensitively dependent on having a certain special kind of law determining the sequence. Moreover, we can also understand why a wide range of perturbations of the system need not disturb the uniform character of the distribution significantly.

8. – The General Applicability of the Calculus of Probability.

Thus far, we have considered only the development of equidistribution in the mean behavior of the angle, θ_n. The calculus of probability deals, however, with properties that are more general than just the mean behavior of a simple event. It deals also with the mean behavior of various combinations of events. Thus, in the case of throws of a die, the theory of probability permits us to estimate the relative frequency with which we shall obtain two sixes in a row, by multiplying the probabilities for each result separately, according to the well known rules that apply for combinations of « independent » events. Moreover, the theory of probability also applies to problems in which the probability distribution is not uniform.

In this section, we shall show in terms of the same sequence treated in Sec. 7 ($\theta_n = K^n\theta_0$) how a determinate law can lead to the validity of the rules for combining probabilities. In the course of the treatment of the above problem, we shall also show how all of our results can be generalized to cases in which the probability distribution is not uniform.

The first problem that we shall consider is that of estimating the mean number of times in which a given angle θ_i is followed by another specified value, θ_j. More complex combinations involving three, four or more angles may evidently also be of interest.

To treat this problem, it is convenient to introduce as an intermediate concept the notion of *conditional probability*. Thus, in our problem, a typical conditional probability is $p_{i_1 i_2}\mathrm{d}\theta_{i_1}\mathrm{d}\theta_{i_2}$ the probability that if the angle after a given collision lies between θ_{i_1} and $\theta_{i_1} + \mathrm{d}\theta_{i_1}$ the angle after the next collision

will lie between θ_{i_2} and $\theta_{i_2} + \mathrm{d}\theta_{i_2}$. It is evident that the probability of the compound event $i_1 i_2$ is then

$$\mathrm{d}P_{i_1 i_2} = p_{i_1} \cdot p_{i_1 i_2} \mathrm{d}\theta_{i_1} \mathrm{d}\theta_{i_2} . \tag{8.1}$$

A special case of considerable interest arises when $p_{i_1 i_2}$ is independent of i_1. In other words, the probability distribution for the second angle does not depend on the first angle. It is easy to see that in this case $p_{i_1 i_2} = p_{i_2}$. Thus the probability of the compound event becomes

$$\mathrm{d}P_{i_1 i_2} = p_{i_1} \cdot p_{i_2} \cdot \mathrm{d}\theta_{i_1} \mathrm{d}\theta_{i_2} \tag{8.2}$$

and we obtain the well-known product rule, which applies for « independent » events. This rule implies that there is no correlation between successive angles. For if there were correlations the probability of obtaining θ_{i_2} would depend on which angle came before.

We may also be interested in the conditional probability if the angle after a given collision lies between θ_{i_1} and $\theta_{i_1} + \mathrm{d}\theta_{i_1}$ the angle after α_1 additional collisions will be between θ_{i_2} and $\theta_{i_2} + \mathrm{d}\theta_{i_2}$. This we denote by

$$p^{\alpha_1}_{i_1 i_2} \mathrm{d}\theta_{i_1} \mathrm{d}\theta_{i_2} .$$

Moreover, we may be interested in combinations involving more events. Thus, we wish to define probability that if after a given collision the angle is between θ_{i_1} and $\theta_{i_1} + \mathrm{d}\theta_{i_1}$ and after α_1 additional collisions between θ_{i_2} and $\theta_{i_2} + \mathrm{d}\theta_{i_2}$ the angle after still another α_2 collisions will be between θ_{i_3} and $\theta_{i_3} + \mathrm{d}\theta_{i_3}$. It is evident that arbitrarily high order conditional probabilities can be defined in this way. The probability of a compound event $(i_1 i_2 \ldots i_K)$ is then

$$\mathrm{d}P^{\alpha_1 \ldots \alpha_{K-1}}_{i_1 i_2 \ldots i_K} = p_{i_1} \cdot p^{\alpha_1}_{i_1 i_2} \ldots p^{\alpha_1 \ldots \alpha_{K-1}}_{i_1 i_2 \ldots i_K} \mathrm{d}\theta_{i_1} \ldots \mathrm{d}\theta_{i_K} . \tag{8.3}$$

A case of great interest is one in which $p_{i_1 i_2}$ is independent of i_1, $p_{i_1 i_2 i_3}$ independent of i_1, i_2, etc. It is readilly verified that for this case

$$p_{i_1 i_2} = p_{i_2} , \quad p_{i_1 i_2 i_3} = p_{i_3} , \quad \ldots , \quad p_{i_1 i_2 \ldots i_K} = p_{i_K} \tag{8.4}$$

and

$$\mathrm{d}P_{i_1 i_2 \ldots i_K} = p_{i_1} p_{i_2} \ldots p_{i_K} \mathrm{d}\theta_{i_1} \ldots \mathrm{d}\theta_{i_K} . \tag{8.5}$$

This is the case of complete absence of correlation up to order K, or of complete « statistical independence ».

We shall now show that with our sequence, $\theta_n = K^n\theta_0$, the conclusions drawn from using probabilities of compound events are valid. To do this, we must deal with two questions. First, do quantities that would correspond to the probabilities of compound events exist (i.e., do the associated relative frequencies approach limits independent of θ_0)? Secondly, if so what are the values of these quantities?

In order to study this problem, we generalize the statement of equi-distribution given in eq. (7.8). First of all, eq. (7.8) applies only for uniform probabilities. One could however define a more general probability, $P(\theta)$ such that

$$f(\theta) = \int_0^{2\pi} P(\theta)\, f(\theta)\, d\theta \Big/ \int_0^{2\pi} P(\theta)\, d\theta = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\theta_n)\,, \tag{8.6}$$

which would hold even if the distribution were not uniform. In order to prove that a probability, $P(\theta)$, really exists for a given type of sequence of angles θ_n, it would be necessary to show that it is possible to find a function such that eq. (8.6) holds independently of θ_0, except possibly for a set of measure zero. In Sec. 7 we have already done this for the sequence, $\theta_n = K^n\theta_0$, in which case it turns out that $P(\theta)$ is a constant. Evidently, however, a more general problem can be attacked in a similar way. It is not difficult to find a simple determinate sequence leading to a non-uniform $P(\theta)$. In this paper, however, we restrict ourselves to merely pointing out that the mathematical framework is already broad enough to permit the treatment of problems involving non-uniform probabilities.

The probabilities of compound events can be treated similarly. To do this, we consider any integrable function of two variables $f(\theta_1, \theta_2)$. Let us recall that in our sequence both θ_n and $\theta_{n+\alpha_1}$ are determined in terms of θ_0. Thus, once θ_n is given, $\theta_{n+\alpha_1}$ is certainly determined. In general, such a determination would lead to correlations (e.g., $\theta_{n+\alpha_1}$ could be equal to $\theta_n+\pi$). We shall see, however, that in our problem, the relation between $\theta_{n+\alpha_1}$ and θ_n fluctuate sufficiently as n changes so that no such correlations exist.

If a probability $P_{\alpha_1}(\theta_1\theta_2)$ of the compound event exists, then we must have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\theta_n, \theta_{n+\alpha_1}) = \int\int P_{\alpha_1}(\theta_1\theta_2)\, f(\theta_1\theta_2)\, d\theta_1 d\theta_2 \Big/ \int\int P_{\alpha_1}(\theta_1\theta_2)\, d\theta_1\, d\theta_2\,. \tag{8.7}$$

Assuming that a suitable function, $P_{\alpha_1}(\theta_1\theta_2)$ exists, then evidently $P_{\alpha_1}(\theta_1\theta_2)$ can be interpreted as the probability of obtaining an angle between θ_1 and $\theta_1 + d\theta_1$ in the n-th step followed by an angle between θ_2 and $\theta_2 + d\theta_2$ in the $(n + \alpha_1)$-th step.

For the special case in which $P(\theta_1\theta_2) = P(\theta_1)P(\theta_2)$ there is then no correlation between θ_n and $\theta_{n+\alpha_1}$, while a still further simplification is obtained when $P(\theta_1)$ and $P(\theta_2)$ are both constant.

To apply eq. (8.7) to our sequence, we shall restrict ourselves to $f(\theta_n, \theta_{n+\alpha_1})$ that can be approximated adequately in terms of a finite but arbitrarily large number of terms in a Fourier series

$$\frac{1}{N}\sum_{n=1}^{N} f(\theta_n, \theta_{n+\alpha_1}) = \frac{1}{N}\sum_{n,l,m} a_{lm} \exp[i(l\theta_n + m\theta_{n+\alpha_1})] . \tag{8.8}$$

As we shall see later such a function is general enough for most applications. We then have (using $\theta_{n+\alpha_1} = K^{\alpha_1}\theta_n$)

$$\frac{1}{N}\sum_{n=1}^{N} f(\theta_n, \theta_{n+\alpha_1}) = \frac{1}{N}\sum_{n,l,m} a_{lm} \exp[i\theta_n(l + m K^{\alpha_1})] . \tag{8.9}$$

We then write $\theta_n = K^n\theta_0$ and $(l+mK^{\alpha_1}) = C_{lm}$ (provided that $C_{lm}\neq 0$). Then replacing θ_0 by $\theta_{lm} = C_{lm}\theta_0$ we get the sum in eq. (8.9)

$$\sum_{l,m} a_{lm} \frac{1}{N}\sum_{n=1}^{N} \exp[iK^n\theta_{lm}] .$$

Since the distribution of angles, $\theta_n = K^n\theta_{lm}$ is uniform, each term in the above double series will vanish (except for a set of θ_{lm} of measure zero).

This conclusion, however, does not apply for $C_{lm} = 0$ and since K was assumed to be an integer in our original proof of the equidistribution, terms with $C_{lm} = 0$ can always appear. We note however that WEYL has shown that the above sum is zero except for a set of measure zero even when K is not rational. To simplify the problem we shall therefore suppose from now on that K is irrational, so that C_{lm} is never zero, unless $l=m=0$. We therefore conclude that (except for a set of θ_0 of measure zero)

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(\theta_n, \theta_{n+\alpha_1}) = a_{00} . \tag{8.10}$$

But if we assume a probability, $P(\theta_1\theta_2)$ constant, then in eq. (8.7) we have

$$\int\int f(\theta_1\,\theta_2)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2 = a_{00} . \tag{8.11}$$

Thus, we have proved that independently of θ_0, the averages of arbitrary functions, $f(\theta_1\theta_2)$ that are expansible in a finite number of terms in a Fourier

series, are the same as those obtained from a probability $P(\theta_1\theta_2) = \text{constant}$. Hence, we have also proved that there is no correlation between θ_n and $\theta_{n+\alpha_1}$.

It is evident that these ideas can be applied to the case of the probabilities of K-fold combinations of events; and a simple extension of the above results shows that

$$P_{\alpha_1 \dots \alpha_{K-1}}(\theta_1 \dots \theta_K) = \text{constant}, \tag{8.12}$$

so that the K-fold correlations vanish, where K is arbitrarily large.

At this point, it should be noted that the set of values of θ_0 for which the distribution of the *simple* events $\varrho(\theta)$, is not uniform, will not in general coincide with the sets θ_{lm} for which the various distributions of *compound* events fail to be uniform. Thus, if θ_0 is one of the values for which $\varrho(\theta)$ is not uniform, then as we have seen, $\theta_{lm} = C_{lm}\theta_0 = (l + mK^{\alpha_1})\theta_0$ will be a value for which the distribution of twofold events corresponding to a separation of α_1, fails to be uniform. Hence, the more kinds of distributions that we take into account and the more terms that we take in the Fourier series, the more values of θ_0 there will be for which some distribution fails to be uniform. Of course, as long as the number of distributions and the number of terms in the Fourier series is finite, we can eventually obtain a number of elements of the sequence as a whole, N, so large that the total measure of all these special angles is as small as we please. Nevertheless, the more different kinds of distributions and the more terms that are used in the various Fourier series, the larger N will have to be in order to make the measure of this special set less than any number, ε, which represents the precision that is significant in the conditions of the problem of interest.

In order to bring out more clearly what we mean by the notion of significant precision, let us first note that in each problem there will in general exist some range of angles, $\Delta\theta$, such that no important properties within the context of interest will depend on a definition of the angle with a precision greater than $\Delta\theta$. We then introduce the so called characteristic function, χ of the set of angles of interest, which is zero outside the set and equal to unity inside the set. The mean value of χ is then equal to the probability that the system will be within the range $\Delta\theta$. Now, to approximate a function such as χ, we need a Fourier series having a number of terms of the order of $1/\Delta\theta$. Thus, we see how the number of terms in the various series is determined.

In order to show what determines the number of distributions of compound events that needs to be taken, we first point out that in any particular problem, only a finite number of combinations of events will in general be important. For example, in an effort to «beat» a game of coin throws, a

player might choose a « system » of plays, such that if he obtains a head, and after α_1 throws, another head, he will then bet on heads for the play coming α_2 throws later still. The vanishing of three fold correlations is enough to guarantee that this systems will not « beat » the game. More general systems could be invented such that he would bet only when the results appeared in some other specified but predetermined way involving any desired number of throws, and as long as the corresponding combination of events has a uniform distribution, he cannot win systematically by any such choice. A different example which is more complex, but which illustrates the same principle comes from statistical mechanics. Thus, the total potential energy of the molecules of a gas is $E = \frac{1}{2}\sum_{ij} V(\boldsymbol{x}_i - \boldsymbol{x}_j)$ where $V(\boldsymbol{x}_i - \boldsymbol{x}_j)$ is the potential energy of interaction of a pair of molecules. The mean potential energy would then evidently depend on the two-fold correlations in the position distribution of the molecules.

From the above example, we see that the combinations of events that are important may depend partly on what is of interest to us, and partly on the physical conditions of the problem.

When the distribution of arbitrary combinations of compound events is uniform so that the distribution possesses the property of statistical independence, (as is true in our sequence of angles) then one can, by the well known techniques of the theory of probability, deduce a number of important results. For example, let us divide the basic circle into two equal parts, denoting by A, a case in which the angle, is in the first part, and B, a case in which it is in the second part. This division provides us with a model of a process leading to results analogous to those in a game of coin throws. Evidently, in the long run, the angles will be uniformly distributed over A and B (except for a set of θ_0 of measure zero). Moreover, a simple calculation shows that various combinations of events (e.g. combinations in which the n-th event is separated from the next by α_r steps) also have a uniform distribution. Thus we reproduce another property of the game of coin throws. (For example, the probability that a head will be followed α_r steps later by a head is the same as that it will be followed by a tail.)

It is clear that we can now calculate the probabilities of compound events in the same way as is done with the theory of probability. Thus the probability of a combination of n events in which the angle is found n_A times in A and n_B times in B (with $n = n_A + n_B$) is just $W = (\frac{1}{2})^n n!/n_A!\, n_B!$ the well-known binomial distribution. If n, n_A and n_B are large, one may obtain the gaussian approximation $W(n, \delta f) \sim \exp[-(n/2)(\delta f)^2]$ where $\delta f = \delta n/n$ and δn is the deviation of n_A/n from its probability (which is $\frac{1}{2}$).

This result means that if we followed the trajectory for any θ_0 (except for a set of measure zero), we would in the long run obtain a gaussian distribution

in $\delta n/n$. Thus, we see how the gaussian distribution which is so typical of the theory of probability, can come out of a determinate sequence (*).

A further step is to deduce Bernoulli's theorem. For as n becomes larger and larger it is evident that one will eventually obtain a negligible number of cases in which δf is greater than some specified number, ε. This means that for a typical trajectory, practically all choice of a sample of n terms will lead to n_A/n very close to $\frac{1}{2}$.

Finally, let us note that these results can be generalized, by showing that they also follow from types of laws intermediate between determinate and probabilistic. This can be done by methods analogous to those appearing in Secs. 6 and 7, but we shall not carry out such a demonstration in detail here.

9. – A Criterion for Randomness.

It can be seen that in the previous section we have justified the calculus of probabilities, as a means of treating the statistical properties developed in the long run by our sequence, $\theta_n - K^n\theta_0$. In doing this, we have also given the elements with which one can obtain a criterion for randomness.

It is our conclusion that the property of statistical independence (i.e., absence of correlations) furnishes a natural and useful criterion for randomness. At this point, we must emphasize, however, that statistical independence with regard to all possible combinations of events will occur only in rather special problems (such as perhaps in the games of coin throws and dice). More generally, we may expect *approximate* and *partial* statistical independence. The degree of randomness may then be approximate, in the sense that the K-fold correlations vanish only in some degree of approximation, and not exactly. It may be partial in that some of the K-fold correlations vanish, but not all of them. As we have pointed out in Sec. 8, the important problem is then, whether the kinds of correlations that have physical significance in a given problem, and those that are of interest to us in certain applications, will be small enough to be neglected. If this is so, we may say that the distribution is «random in the context under discussion».

As a justification for our proposed criterion for randomness, in terms of absence of correlations, we first point out that this is effectively the practical criterion *most frequently used* in applications in physics, (e.g., statistical mechanics) (†) as well as in other applications of statistics (e.g., problems

(*) The gaussian approximation has already been obtained from determinate sequences in other ways. See, for example, M. KATZ and H. STEINHAUS: *Studia Mathematica*, **6**, 59 (1936).

(†) In Section **10**, we shall give an example of how this definition of randomness is used in statistical mechanics, in connection with the development of «random phases».

of distributions of various kinds of members in a certain population). Secondly, we point out that it contains what is really essential in the definition given by VON MISES (*). For if a sequence has the property of statistical independence of its members, then as we have pointed in Sec. 8, the gaussian law follows, so that most samples will tend to give good representations to the relative frequency distribution in the sequence as a whole.

The criterion that we have given for randomness avoids the problems appearing in the definition given by VON MISES. First of all, we do not define randomness in terms of irregularity or lawlessness and thus, we do not become involved in the very probably unsolvable problem of defining what irregularity and lawlessness mean. Instead, we regard randomness as a special manifestation of law, a manifestation which can appear however only in the context of a statistical aggregate and one of whose characteristic aspects is statistical independence. In this regard, we adopt a position similar to that of WALD (+) and various other members (*) of the school of VON MISES, who adopt as a criterion for randomness a certain class of laws suggested by the impossibility of « beating » a gambling game. Our criterion differs, however, from theirs in two aspects: (*A*) It is much easier to apply in practice in a wide range of statistical problems, and indeed corresponds to just the practical definition that is most frequently used. (*B*) The criteria for randomness are not chosen completely a-priori or by the simple transfer to all possible problems of the criteria that may be significant in a given problem (e.g., gambling games). Instead, the criteria are determined either by the actual physical situation or by the statistical problems that we wish to treat, because the kinds of correlations that are significant may depend on the context in which the theory is applied.

With regard to the question of irregularity or law-lessness, it appears to us that this is not an essential feature of the problem. For example, in the case of throws of a coin, regular orders, such as that of ten heads in a row will certainly be quite rare. But any other *specified* order, however irregular it may seem to us, will be equally rare. Thus, the rarity of « regularity » is not so much a property of the sequence itself as of our own tendency to pick out a few special orders and call them « regular », while the overwhelming majority of orders are relegated to the very broad category of « irregular ». Hence any distribution that fluctuates over a wide range of possible orders will inevitably seem irregular to us most of the time.

Before ending this section, we shall make a few remarks on the relationship between randomness and chance. Now, as we have seen, randomness

(*) See Section 4.
(+) See Section 4.

is a *statistical* property of a given system. It may evidently come from the effects of external disturbances that already have a degree of randomness in them, and it may also arise from purely internal causes (as happened with our sequence $\theta_n \quad K^n\theta_0$); but most generally it arises because of some combination of these two kinds of causes. The assumption that statistical theories must always be restricted to the calculation of one kind of probability in terms of another is then equivalent to neglecting the possibility that a part of the randomness, can arise internally in a system, without the action of some external causes that already have in them some degree of randomness.

On the other hand, as we saw in Sec. 2, chance refers to a relationship in which the events in a given context depend contingently on « independent » fluctuating causal factors lying outside the context of interest. Thus, while a random distribution can develop of its own accord in a given context, chance can by its very definition refer only to the effects of causes that lie outside the context under discussion.

There is, nevertheless, a close relationship between randomness and chance. For in those cases of chance fluctuations that are most typical (e.g., coin throws), the external contingencies (e.g., the motions of the hand that throws the coin) on which the results depend, have a range of fluctuation within which the distribution is effectively random. Such cases, correspond to the category of « pure chance » discussed in Sec. 2. There may, however, exist a more general kind of dependence in fluctuating causal factors (*) that are not entirely random in their distribution. This dependence still leads to a kind of chance, but it is not pure chance. In such a case, the calculus of probabilities evidently would be at best an approximation. Here, then, we have an additional example of the general statistical problem; for it is evidently of interest in such cases to study the question of how good is the degree of approximation provided by the theory of probability, as well as perhaps to seek a more general calculus that might apply even when we do not have pure chance.

10. – Further Illustration in Terms of Random Phases.

In this section, we shall apply some of the ideas suggested in the previous sections to a problem that has widespread application in statistical mechanics: namely that of the development of random phases in a system of oscillators or rotators.

The simplest example is the problem of random phases with a system of harmonic oscillators. Let x_i be the coordinate of a given oscillator, y_i the

(*) This possibility has also been discussed in Section 2.

corresponding canonically conjugate momentum. The equations of motion of such a set of oscillators (which are, in general, coupled), can be derived from the hamiltonian

$$H = \tfrac{1}{2} \sum_{i,j} (a_{ij} y_i y_j + b_{ij} x_i y_j + c_{ij} x_i x_j) . \tag{10.1}$$

As is well known, there exists a linear transformation, to the normal coordinates ([11]), q_i and to the conjugate momenta, p_i, such that the hamiltonian reduces to a sum of squares, each equivalent to a separate harmonic oscillator, not coupled to the others.

$$H = \tfrac{1}{2} \sum_{i} (A_i p_i^2 + B_i q_i^2) . \tag{10.2}$$

The equations of motion ($\dot{q}_i = \partial H / \partial p_i$, $\dot{p}_i = - \partial H / \partial q_i$) then lead to

$$\ddot{q}_i + (B_i/A_i) q_i = 0 , \qquad \ddot{p}_i + (B_i/A_i) p_i = 0 \tag{10.3}$$

so that each oscillator has angular frequency $\omega_i \quad (B_i/A_i)^{\frac{1}{2}}$. The solutions can then conveniently be expressed in the complex form

$$q_i + i p_i = K_i \exp [i \omega_i t] , \tag{10.4}$$

where K_i is an arbitrary complex number, which can be written as

$$K_i^2 = J_i \exp [2 i \varphi_{0i}] , \tag{10.5}$$

$$q_i + i p_i = (J_i)^{\frac{1}{2}} \exp [i \omega_i t + \varphi_{0i}] . \tag{10.6}$$

Evidently J_i is proportional to the energy of the j-th oscillator and φ_{0i} is a constant of the motion corresponding to the initial value of the phase angle of the j-th oscillator.

As a function of the time, the phase angle is then given by

$$\varphi_i = \varphi_{0i} + \omega_i t . \tag{10.7}$$

As we shall show in this section, if the ω_i are incommensurable (as they usually are in most applications) the phases, φ_i will in time become random relative to each other.

The problem of random phases arises in a great many important problems

([11]) H. G. GOLDSTEIN: *Classical Mechanics* (1952), Chap. 10.

in statistical mechanics, e.g., thermal vibration of a solid, the thermal excitation of rotating molecules, motions which can be solved in terms of angle and action variables and most important of all, in the foundations of quantum statistical mechanics ([12]). The application of the notion of random phases in the foundations of quantum statistical mechanics, we shall discuss in more detail at the end of this section.

In order to illustrate in a simple way just how the phases become random (*), let us begin by considering the case of two variables. Now, in the problem that we are treating here, we have the new feature, not present in the treatment given in sections 7 and 8, that instead of a sequence of discrete values, we have to deal with a continuous function of the time t. For this reason, we shall replace the sum over the N terms of the sequence appearing in eq. (8.7) by an integral over the time, representing the mean value of an arbitrary function, $f(\varphi_1(t), \varphi_2(t))$. (The use of the continuous time average in this case corresponds to the actual practice in the definition of physical quantities in statistical mechanics ([13])). We then obtain as a criterion for a uniform distribution that the following relationship must hold independently of initial conditions (except possibly for a set of measure zero):

$$\lim_{T\to\infty}(1/T)\int_0^T f(\varphi_1(t),\varphi_2(t))\,\mathrm{d}t = \int_0^{2\pi}\int_0^{2\pi} f(\varphi_1,\varphi_2)\,\mathrm{d}\varphi_1\,\mathrm{d}\varphi_2\,. \tag{10.8}$$

The limits on φ_1, φ_2 are taken from zero to 2π, because as can be seen from (10.6), the system comes to the same physical state when φ_i or φ_2 change by 2π. This means also that all physically significant functions of φ_1 and φ_2 can be expressed as periodic functions with period 2π or that

$$f(\varphi_1,\varphi_2) = \sum_{i,j} f_{ij}\exp\left[i(n_i\varphi_1 + n_j\varphi_2)\right], \tag{10.9}$$

n_i and n_j being integers.

Now, writing $\varphi_1 = \varphi_{01} + \omega_1 t$, $\varphi_2 = \varphi_{02} + \omega_2 t$ we obtain

$$f(\varphi_1\varphi_2) = \sum_{i,j} f_{ij}\exp\left[i(n_i\varphi_{01} + n_j\varphi_{02}) + i(n_i\omega_1 + n_j\omega_2)t\right]. \tag{10.10}$$

Inserting (10.10) in (10.8), we note that if ω_1 and ω_2 are not commensur-

([12]) See, for example, R. C. Tolman: *The Principles of Statistical Mechanics*, (Oxford, 1950), Chap. IX.

(*) The development of random phases is already well known (see, for example, reference ([9])). However, we shall give a brief sketch of a proof here which is convenient for illustrating the concepts involved.

([13]) See, for example, reference ([6]).

able every term on the left hand except f_{00} will integrate out to zero as the time T approaches infinity. On the right hand side, we see from (10.9) that likewise only f_{00} will survive integration over φ_1 and φ_2. Thus we demonstrate that the criterion for equidistribution is satisfied independently of initial conditions (*) (note that in this problem, there is no exception at all, not even of a set of measure zero). We conclude then that there exists a uniform probability density that the phase point will occupy any specified region of the configuration space of φ_1, φ_2.

For a more general case, in which there does not exist a uniform distribution, eq. (10.8) should be replaced by

$$\lim_{T\to\infty} (1/T)\int_0^T f(\varphi_1(t), \varphi_2(t))\,\mathrm{d}t = \int_0^{2\pi}\int_0^{2\pi} P(\varphi_1, \varphi_2) f(\varphi_1, \varphi_2)\,\mathrm{d}\varphi_1\,\mathrm{d}\varphi_2\,, \tag{10.11}$$

where $P(\varphi_1, \varphi_2)$ is the probability density in the configuration space of φ_1, φ_2 and where the equation must hold independently of initial conditions (except for a possible set of measure zero), if the motion is to be such as to permit the definition of a probability.

To show in more detail how the equidistribution comes about in the case that we have treated, let us consider the φ_1, φ_2 plane as shown in Fig. 5, where φ_1, φ_2 go only over their physically significant range of 2π. Consider an initial point, P, with $\varphi_1 = \varphi_{01}$ and $\varphi_2 = \varphi_{02}$. Then with the passage of time, the phase point will move until it finally reaches the point, Q, where one of the variables, φ_1 or φ_2 is equal to 2π (in the diagram, it is φ_2 that is equal to 2π). At this point, P must decrease abruptly by 2π and jump to the point, Q'. Then the point moves until it reaches Q'', then jumps to Q''' moves to Q^{IV}, etc. The theorem of equidistribution then shows that in time the phase point will fill the above region equidensely. Evidently this can happen only if ω_1 and ω_2 are incommensurable, for if they are commensurable, the phase point will describe a closed curve.

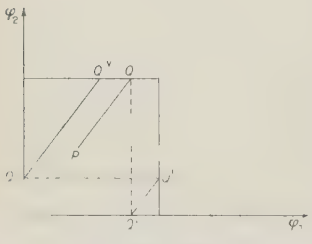


Fig. 5.

In the space of the p_i and the q_i we obtain the well-known Lissajous figures, which fill the space densely, but not in general equi-densely.

The generalization to any number of variables is quite straight-forward; and we need only state the result that if all the ω are incommensurable, then

(*) This result is essentially a proof of the quasi-ergodic theorem for the special problem that we are treating. (See section 6).

the distribution will have a uniform probability density in the configuration space $(\varphi_1 \dots \varphi_N)$. Eqs. (10.8) and (10.11) then represent a statement of the fact that the time average behavior of a single system can be represented by a suitable average over angles of an ensemble of systems, taken at a single instant of time, t. Thus, the use of statistical ensembles (which are essentially the same as the collectives of VON MISES) to represent the statistical properties of a single system is justified. The uniformity of the distribution in the configuration space evidently implies that $P(\varphi_1, \varphi_2 \dots \varphi_N) = P(\varphi_1)P(\varphi_2) \dots P(\varphi_N)$ so that the phase angles of the different oscillators are uncorrelated, and therefore statistically independent. Physicists have gradually come to refer to a system of variables that behaves in this way as « random ». (Thus, we obtain a justification of the definition of randomness suggested in Sec. 9, at least as far as applications in physics are concerned).

As we have already pointed out, the problem of random phases has a wide application in statistical mechanics. This problem is particularly important in quantum theory, because it enters as an essential part of the foundations of quantum statistical mechanics. To see how this comes about, we note that, as is well known, the quantum state of a system is defined by its wave function† $\psi_E(\boldsymbol{x}_1 \dots \boldsymbol{x}_n)$ where $\boldsymbol{x}_1 \dots \boldsymbol{x}_n$ are the coordinates of the system. An arbitrary wave function can be expressed in terms of the series of eigenfunctions $\psi_{E_n}(\boldsymbol{x}_1 \dots \boldsymbol{x}_n)$ of the Hamiltonian operator, corresponding to eigenvalues, E_n. Thus,

$$\psi = \sum_n C_{E_n} \psi_{E_n}(\boldsymbol{x}_1 \dots \boldsymbol{x}_n) \exp[-iE_n t/\hbar]\,. \tag{10.12}$$

(We are assuming the system to be in a bounded region of space, so that the eigen-values are discrete, and we are assuming that there is no degeneracy. The case of degeneracy requires a more extensive treatment, which we shall not give here.)

We then write

$$C_{E_n} = a_n \exp[i\varphi_{0n}]\,, \tag{10.13}$$

obtaining

$$\psi = \sum_n a_n \psi_{E_n} \exp[i\varphi_n]\,, \tag{10.14}$$

with

$$\varphi_n = \varphi_{0n} - E_n t/\hbar\,. \tag{10.15}$$

Comparing (10.14) and (10.15) with (10.6) and (10.7) we see that formally

speaking, the wave function acts like an infinite collection of harmonic oscillators, one for each eigenfunction of the hamiltonian. This is not surprising, since the Schrödinger equation is linear in ψ and as is well known a wave field satisfying a linear partial differential equation can be expressed in a series of terms each oscillating with its own characteristic frequency.

Now we note that in general, the angular frequencies $\omega_n = E_n t/\hbar$ of oscillation of the wave function in different stationary states, will not be commensurable (*). In the long run, we shall therefore have a uniform distribution in the configuration space of the phases, and the phases will be random relative to each other.

Finally, we remark that the problem of random phases, is formally of the same type as the one treated by POINCARÉ in connection with the distribution of the phase angles of a swarm of asteroids in neighbouring orbits (see Section 8).

11. – Conclusion.

In this paper, we have shown that in a large number of statistical problems arising in physics, it is either advantageous or necessary to go outside the framework of concepts that appear in the theory of probability. Among such problems are the following:

(1) The calculation of the large-scale statistical and thermodynamic properties of an aggregate containing a large numbers of molecules.

(2) The treatment of non-equilibrium states of such an aggregate.

(3) The demonstration that a given kind of sequence contains statistical relationships that permit its long-run behavior to be treated by means of the calculus of probabilities.

(4) The estimate of error in the theory of probability.

(5) The treatment of approximate and partial randomness.

(6) The treatment of chance fluctuations originating in contingencies that are not totally independent of what happens in the context under discussion.

An inspection of the above list would suggest that the problems of statistical physics in which the theory of probability does not apply are perhaps

(*) They are commensurable for the one-dimensional harmonic oscillator, but not in most other general problems.

even more important than those in which it does apply. A systematic attack in statistical problems going outside the scope of the theory of probability, coupled with an effort to develop new formal systems that are more general than that of the calculus of probabilities, would therefore be likely to produce fruitful results.

* * *

We wish to thank Professor MARIO SCHÖNBERG, Dr. MARIO BUNGE and J. A. MEYER for many stimulating and instructive discussions, which played a very important role in the clarification of our ideas.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

On Large Scintillation Counters.

D. Brini, L. Peli, O. Rimondi and P. Veronesi

Istituto di Fisica dell'Università - Bologna
Istituto Nazionale di Fisica Nucleare - Gruppo di Bologna

(ricevuto il 23 Marzo 1955)

Contents: 1. Introduction. – 2. Generality. - 3. On the Uniformity of a Counter with Transparent Walls. 3·1. Geometrical Considerations. 3·2. Absorption Influence. 3·3. Influence of the P.M. Photocathode Sensitivity. 3·4. Influence of Light Distribution on the Photosensitive Surface. – 4. Uniformity of a Counter with Reflecting Walls. – 5. Uniformity of a Counter with Diffusing Walls. - 6. Experimental Proofs. 7. Light Collection in Time. – 8. Conclusions. – Appendices I, II, III.

1. – Introduction.

The difficulties in connexion with the construction of large proportional scintillation counters are well known. It can be plainly said, that it is no longer useful to speak of proportionality, when the three counter dimensions become excessively great. If only two, or one, of the three dimensions are very great, one can then still speak of proportionality, provided that one uses a sufficiently high number of photomultipliers.

We wished to make a detailed study of the causes of disuniformity in counters measuring some decimetres, for which the problem of the realization of proportional counters can still reasonably be put.

When it is necessary to construct counters of this type, we find ourselves faced with a very wide possibility of choice. In fact, a choice can be made of transparent, reflecting or diffusing walls; a greater or smaller number of P.M.'s can be used, which can be arranged to look the scintillator in the most varied manners; the scintillator can be chosen from a large quantity of substances which are suitable for this purpose and, finally, if the experiment does not expressly require a special shape, even the form of the counter can be

arbitrary. It is somewhat difficult to establish *a priori* what can be, from the point of view of economy, the most convenient combination of all these elements.

In this work we wished to put in evidence the possible criteria which can point the way to a resolution of each particular problem.

We conducted our analysis beginning with a particular examination of a transparent-walled parallelepiped-shaped counter, establishing in this case some criteria which permit, from the point of view of designing, an estimation of the uniformity which can be obtained with this arrangement. The extension of the analysis to other shapes and wall types, which was carried out by direct comparison with the first one studied, allowed us to arrive at some conclusions which may be of use when a counter of the above-mentioned dimensions is planned.

The research on the temporal behaviour of light collection was also carried out with the same criterium. That permitted us to establish the inferior limit of the resolving time of these counters which does not influence the uniformity of the response.

2. – Generality.

We shall say that a scintillation counter has a uniform response when the electric charge collected at the P.M. anode is proportional to the light emitted in scintillation and independent of the position at which the emission took place. Such a counter, for the sake of brevity, will be called uniform. This condition is necessary, but not sufficient, for it to be proportional (*).

The realization of a uniform counter shows itself to be more complex as the dimensions of the said counter are greater.

The following factors influence the uniformity:

1) the type of walls (diffusing, transparent, reflecting), the shape of the counter and the position in which the P.M.'s are placed;

2) the absorption of the walls and the scintillator;

3) the disuniformity of the photosensitive surface (+).

The first two factors have an influence on uniformity in light collection, which is realized when the relation between the light which falls on the photosensi-

(*) Proportionality implies considerations on the energy loss in the counter and, therefore, on the type of particle to be counted and on the nature of the phosphor.

(+) This expression is intended to indicate:

a) in the case of a single P.M. the different sensitivity from point to point of the photocathode;

b) in the case of more than one P.M. the different gain also.

tive surface and that which is emitted is independent of the position at which scintillation takes place.

To understand the rôle played by the third factor it is necessary to bear in mind that light is not distributed uniformly on a photosensitive surface. In fact, the quantities of incident total light being equal, the distribution depends on the position at which scintillation takes place.

3. – On the Uniformity of a Counter with Transparent Walls.

3·1. *Geometrical Considerations.* – In a transparent-walled counter one part of the light emitted in scintillation remains entrapped and another escapes. In order to get uniform collection, and leaving out for the moment the question of absorption, it is necessary (but not sufficient) for the percentage of entrapped light to be independent of the position at which scintillation takes place. Now, spherically [1] and cylindrically [2] shaped counters do not satisfy such a condition. The one geometrical shape which realizes uniform entrapping is that of a parallelepiped [1,3]. In fact supposing that the material constituting the parallelepiped to be homogeneous and given the limit angle of this medium in respect to the air as γ, it can be seen that there exist six 2γ aperture cones through which light escapes (*). It is easy to estimate the percentage of entrapped light [1] once the refractive index n has been given.

Simple geometrical considerations show clearly that it is necessary, if all the entrapped light is to be collected by the photosensitive surface, for the latter to cover *at least one entire face* of the parallelepiped.

Every other arrangement of the photosensitive surface which does not satisfy such a condition, always causes the loss of one part of the entrapped light, a part which depends on the position of scintillation. It is necessary to remark, however, that, with an entire face which is photosensitive, not only the entrapped, but also the direct, light is collected. In order for this contribution to be independent of the scintillation position, it is necessary that scintillation should never take place at a distance lower than $l = d \cdot \mathrm{tg}\,\gamma$ (see Fig. 1) from the photosensitive face, where d is the diagonal thereof. In other words, the photosensitive face must be placed at a distance l from the

[1] W. A. Shurcliff and R. C. Jones: *Journ. Opt. Soc. Am.*, **39**, 912 (1949).

[2] D. Brini, L. Peli, O. Rimondi and P. Veronesi: *Nuovo Cimento*, **11**, 655 (1954).

[3] R. H. Gillette: *Rev. Scient. Instr.*, **21**, 294 (1950).

(*) Rigorously [4], a small percentage of this light is reflected and partially polarized. On an average these rays undergo other reflections in order to reach the photosensitive surface, through which, even setting aside scintillator absorption, they are in practice rapidly extinguished.

[4] F. A. Jenkins and H. E. White: *Fundamentals of Optics* (London, 1953), (Chapter 28).

scintillator. Naturally this space must be filled with material having the same refractive index as the scintillator. Such optical connexion is normally called pipe light. For successive calculations, the pipe light will be supposed as a prolongation of the parallelepiped (Fig. 1), which is not generally the case in practice.

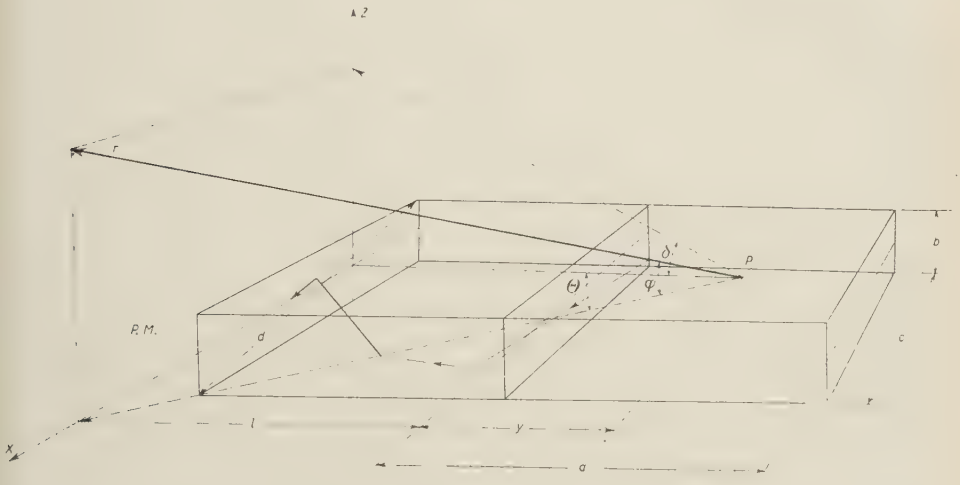


Fig. 1.

From now on, then, we shall make a distinction between the collection surface and the photosensitive one, which are the two opposite surfaces of the pipe light, distant l from one another, and parallel.

In a counter so made, and continuing to leave out the question of absorption, and under the condition $\gamma < 45^\circ$ $(n > \sqrt{2})$ the percentage of light collected will appear as

$$R = \tfrac{1}{2}(5 \cos \gamma - 3) \cdot 100\% .$$

For the refractive index $n = 1.5$, such a percentage, which will hereafter be called the collection yield, appears as $R \simeq 36\%$. With two opposite photosensitive faces a yield of $R \simeq 49\%$ is obtained.

The same yield could be obtained with one photosensitive face only making the opposite face reflecting, provided that the reflection coefficient be one.

It is then evident that with an increase in the number of photosensitive faces the yield only increases (the question of absorption is as always set aside). Obviously there is very little practical advantage to be gained from considering a counter with more than two entire collection faces.

From what has been said above it is apparent that in order to have a transparent-walled counter with light collection and good yield, it is necessary

that following essential conditions be satisfied:

a) parallelepiped shape;

b) at least one entire collection face;

c) pipe light with a minimum length of $l = d \cdot \mathrm{tg}\,\gamma$.

In practice, except in the case of a phosphor-plastic mixture, the scintillator is always placed inside a container, the walls of which will be either diffusing, reflecting or transparent. We shall examine the last case first making the hypothesis (which in practice can be done with a good approximation) that the refractive index of the material which constitutes the container is identical to that of the phosphor. Under this condition it is the container which must satisfy the required geometrical conditions. That means that the parallelepiped is limited by the external faces of the container, while the scintillator inside it can have any form whatsoever. Nevertheless, if this condition is not required by particular experimental exigencies, it is convenient for the scintillator also to have a parallelepiped shape . This implies that the container walls be made of plain parallel surfaces.

For the discussion of the causes of disuniformity 2) and 3) we shall continue to refer to a counter which satisfies the preceding conditions.

3·2. *Absorption influence.* – When the dimensions of the counter (and, therefore, the ranges of light) are considerable, light absorption on the part of the walls and the scintillators plays a very important rôle in regard to uniformity. Such absorption, in addition to having an influence on uniformity, obviously influences the yield also (together with the absorption due to the pipe light). In designing a big counter it is, therefore, of great importance to choose the most suitable materials to be used, with the aim of reducing to a minimum the disuniformity due to absorption, compatible with the dimensions of the counter itself. It is obvious that, in order to make exact calculations in this connexion, it is necessary to bear in mind:

a) the scintillator emission spectrum;

b) the scintillator absorption spectrum;

c) the absorption spectrum of the material of which the walls and pipe light are made.

Unfortunately in practice, exact information on these data is not always available. Insofar as regards the emission spectrum, it is known that the relative measurements are neither simple nor very exact ([5-7]).

([5]) H. KALMANN and M. FURST: *Phys. Rev.*, **81**, 853 (1951).
([6]) F. B. HARRISON: *Nucleonics*, **10**, n. 6, 40 (1952).
([7]) R. C. SANGSTER: *M.I.T. Technical Report* 55 (January 1, 1952).

The absorption spectra can, on the contrary, be sufficiently well measured abstracting from the indetermination due to the purity of the phosphor.

To estimate the absorption by the walls, scintillator and pipe light of the light emitted in scintillation, let us suppose, as we mentioned before, the refractive indices of the various materials (scintillator, walls and pipe light) to be equal.

Let us again suppose that scintillation is in a point region. Let us consider Fig. 1 and indicate with:

— a, b, c, the edges of the parallelepiped which has, as a collection surface, a surface cb;
— y, the distance of scintillation P from the collection surface (this is the only co-ordinate of P which is of interest for these considerations in this particular geometrical form of the counter);
— r, the generic ray escaping from P;
— x_1 a light range in a wall ac (Fig. 2);
— s_1 the thickness of the horizontal walls;
— x_2 a light range in a wall ab (Fig. 2);
— s_2 the thickness of the vertical walls, parallel to y;
— x_3 the light range in the pipe light;
— $l = d \cdot \mathrm{tg}\,\gamma$ the length of the pipe light;
— x_4 the light range in the scintillator;
— $\alpha_s(\lambda)$ the scintillator absorption coefficient;
— $\alpha_w(\lambda)$ the wall absorption coefficient;
— $\alpha_p(\lambda)$ the pipe light absorption coefficient;
— N_1 the number of reflections on the horizontal walls;
— N_2 the number of reflections on the vertical walls, parallel to y.

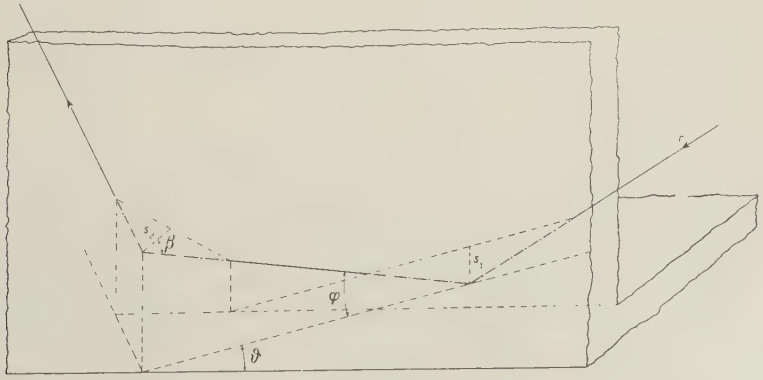


Fig. 2.

Let us suppose that, in a scintillation, a number of photons $\mathcal{N}_0(\lambda)$ of a wavelength between λ and $\lambda + \mathrm{d}\lambda$ are emitted isotropically. The number of

photons emitted in the solid elementary angle $d\omega = \cos\varphi\, d\varphi\, d\theta$ which arrives on the photosensitive face is given by

$$(1) \qquad d\mathcal{N}(\lambda) = \frac{\mathcal{N}_0(\lambda)}{4\pi} \exp\left[-\alpha_W(\lambda)(N_1 x_1 + N_2 x_2) - \alpha_p(\lambda) x_3 - x_s(\lambda) x_4\right] d\omega\, d\lambda\,.$$

An approximate calculation of N_1 and N_2 can be made by considering the projections of the generic ray r escaping from scintillation, one on a vertical plane and the other on a horizontal plane (Fig. 1). We then obtain

$$N_1 = \frac{y}{b} \operatorname{tg} \delta = \frac{y}{b} \frac{\operatorname{tg} \varphi}{\cos\theta}\,; \qquad N_2 = \frac{y}{c} \operatorname{tg} \theta\,.$$

N_1 and N_2 will evidently have to be integers. Assuming them as given by the two preceding expressions, we make, in effect, an overestimation of the absorption due to the walls.

From Fig. 2 it can immediately be deduced, respectively, that:

$$x_1 = \frac{2s_1}{\sin\varphi}\,; \qquad x_2 = \frac{2s_2}{\cos\beta} = \frac{2s_2}{\cos\varphi \sin\theta}\,.$$

Analogously it can be found that:

$$x_3 = \frac{l}{\cos\varphi \cos\theta}\,; \qquad x_4 = \frac{y}{\cos\varphi \cos\theta} - (N_1 x_1 + N_2 x_2)\,.$$

Then (1) becomes

$$(2) \qquad d\mathcal{N}(A, \theta, \varphi) = \frac{\mathcal{N}_0(\lambda)}{4\pi} \exp\left[-A(y, \lambda)/\cos\varphi \cos\theta\right] \cos\varphi\, d\varphi\, d\theta\, d\lambda\,,$$

where it is

$$A(y, \lambda) = y\left[\alpha_s(\lambda) + 2\{\alpha_w(\lambda) - \alpha_s(\lambda)\}\left(\frac{s_1}{b} + \frac{s_2}{c}\right)\right] + l\alpha_p(\lambda)\,.$$

We shall carry out the integration of (2) in respect only to θ and φ, considering an interval $\Delta\lambda$ of λ within which, practically speaking, the quantities dependent on it can be considered constant. The relative quantity of light $\left(\text{given by } \Delta R(A) = \int_{\varphi}\int_{\theta} \frac{d\mathcal{N}(A, \theta, \varphi)}{\mathcal{N}_0(\lambda)}\right)$ received by the photonsensitive surface can be considered split in two contributions ΔR_t and ΔR_b:

1) ΔR_t represents the quantity of the light emitted by the source which reaches the semispace turned towards the photosensitive face in respect to a plane parallel to this;

2) ΔR_b is the analogous contribution due to the light emitted by the opposite part of the plane considered, which, through successive reflections, can reach the photosensitive face itself.

Strictly speaking it would be necessary to bear in mind, when calculating this contribution, the absorption due to the one range in the wall opposite to that of collection. That is expressed in a prolongation of the length l of the pipe light by a quantity (in the case of $\alpha_w = \alpha_p$) equal to $2s_2$.

Our calculations were carried out without paying attention to $2s_2$ in respect to l, a condition which, in practice, it is opportune to realize.

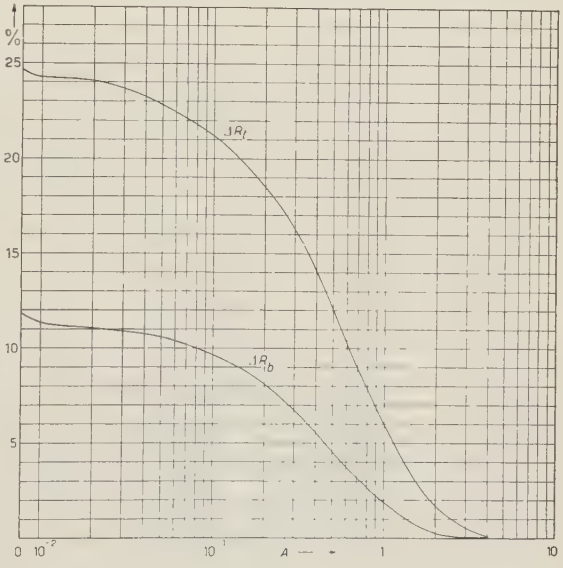


Fig. 3.

The detailed calculations will be reported in Appendix I. These calculations allow for the construction of the two curves ΔR_t and ΔR_b (Fig. 3) which give us, respectively, the yields due to the two contributions 1) and 2) as functions of A. Simple geometrical considerations show that, for a given value of y, contribution 1) corresponds to a certain value $A(y, \lambda)$ and contribution 2) to a value of $A(2a - y, \lambda)$. For a generic scintillation taking place at a distance y, these curves allow the estimating of the relative contribuitons ΔR_t and ΔR_b, which refer to the various intervals $\Delta\lambda$ which satisfy the conditions appointed in the beginning, in which the emission spectrum can be considered subdivided. The total contribution ΔR, for an interval, is expressed by:

$$\Delta R = \Delta R_t + \Delta R_b ,$$

while the total contribution due to all the spectrum intervals considered will evidently be given by

$$\sum \Delta R .$$

In the case in which the photosensitive faces are two entire opposites, the same considerations unfolded above are valid, except that it is necessary to take into consideration only the curve ΔR_t in correspondence to the two values $A(y, \lambda)$ and $A(a - y, \lambda)$. The relative contribution of light collected then,

corresponding to the interval $\Delta\lambda$, will be:

$$\Delta R = \Delta R_t + \Delta R'_t ,$$

where ΔR_t is calculated for $A(y, \lambda)$, $\Delta R'_t$ is calculated for $A(a - y, \lambda)$.

This arrangement permits an improvement both in the yield as in the uniformity of collection.

We might think of reaching the same results by substituting a photosensitive surface by a reflecting one. In this case, maintaining the usual symbolism, we should obtain:

$$\Delta R = \Delta R_t + k\Delta R''_t ,$$

where ΔR_t has its usual meaning, and $\Delta R''_t$ corresponds to a $A(2a - y, \lambda)$ and k is a factor < 1 which represents the reflection coefficient. In this case the yield and collection uniformity are midway between the two cases treated in precedence. The results obtained relative to a counter studied experimentally will be referred to hereinafter; these will help to clarify the conclusions drawn above.

3·3. *Influence of the P.M. Photocathode Sensitivity.* – The preceding remarks refer to the quantity of light collected by the photosensitive surface independently of the spectral sensitivity characteristic of the P.M. It is evident, however, that, given a certain emission spectrum, the collected light spectrum will be different with the varying of y because of the selective absorption of the scintillator, walls and pipe light. Since the P.M.'s too present selective sensitivity for the various wavelengths, this is expressed by the fact that the number of photons being equal, there is a different charge response. Using the P.M. sensivity curves and the results of § 2 which permit to express the relative quantity of light collected in the spectral interval λ and $\lambda + \mathrm{d}\lambda$, the charge accumulated at the anode by the effect of this contribution will be:

$$\Delta Q(\lambda) = g \cdot \Delta R \cdot \mathcal{T}(\lambda) \ (*) .$$

(*) Assuming Birks' [8] notations g corresponds to the product

$$C_{PE} f(\nu_P) \cdot (CR)^m$$

where the first factor is the photoelectric conversion efficiency of the cathode and the second is the overall gain of the tube which is expressed by C (collection efficiency factor of the photoelectrons by the dynode system) R (electron multiplication factor) m (number of the dynodes). $C_{PE} f(\nu_P)$ is a quantity which is variable from point to point on the photocathode surface [9] wherefore, in a very strict estimate, it would be necessary to bear also this in mind. We shall see later on how the calculation can be made.

[8] J. B. Birks: *Scintillation Counters* (London, 1953), Chapter 2.

[9] F. H. Marchall, J. W. Coltmann and A. I. Bennet: *Rev. Scient. Instr.*, 19 744 (1948).

3·4. *Influence of Light Distribution on the Photosensitive Surface.* – Another cause of disuniformity is to be found in the different values of the ratio $Q/\mathcal{N}=g$ (total charge of the anode/number of photons fallen on the photocathode) for the different P.M.'s which look at the counter, $\mathcal{N}$ being equal. The shape of the counter being known, its dimensions, type of walls, etc. (characteristics which have already been chosen for the best uniformity in light collection), it is now therefore a problem of seeing in which way the use of any number of P.M.'s influences the uniformity of the electric signal with the varying of the position at which scintillation takes place. Should the n P.M. all have the same value of the $Q/\mathcal{N} = g$ ratio, it is obvious that there would be no cause for disuniformity. If the parameter g of each tube were different, but the quantity of light which falls on each one of them remain constant (even if differing from tube to tube) with the varying of the scintillation position, there would not be disuniformity even in this case. It should be a question, on the contrary, of disuniformity, because the values of g are different among themselves, and the quantity of light received separately by the various tubes varies with the varying of the scintillation position.

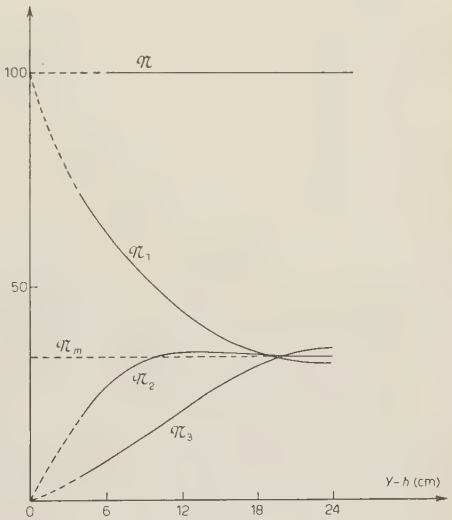


Fig. 4.

These considerations can be determined quantitatively. Let $\mathcal{N}$ be the total quantity of light received from the photosensitive face and $\mathcal{N}_i$ that received from the i-th tube. Let us suppose as given the behaviour of all the $\mathcal{N}_i$'s in function of the scintillation position P (Fig. 4). (The figure reports the case of $n = 3$). Whatever their value may be, a point-by-point deviation $\Delta\mathcal{N}_i$ from the mean value $\mathcal{N}_m = \mathcal{N}/3$ can be determined. If the counter is optically uniform, it is $\mathcal{N} = \text{cost}$, i.e. the light received from the photosensitive face is the same whatever P may be. Bearing in mind the sign of $\Delta\mathcal{N}_i$, it follows that $\sum \Delta\mathcal{N}_i = 0$. The charge contribution for each tube is

$$Q_i = g_i \mathcal{N}_i \,,$$

where Q_i indicates the value of the charge obtained at the output of the i-th P.M. The last relation can also be written

$$Q_i = g_i \mathcal{N}_m + g_i \Delta\mathcal{N}_i \,.$$

Summing up the n tube contributions, we get

$$Q_{i\mathfrak{T}} = g_i \mathcal{Q}_m + g_i \Delta\mathcal{Q}_1 + g_2 \mathcal{Q}_m + g_2 \Delta\mathcal{Q}_2 + \ldots + g_n \mathcal{Q}_m + g_n \Delta\mathcal{Q}_n = \\ = \sum_i g_i \mathcal{Q}_m + \sum_i g_i \Delta\mathcal{Q}_i .$$

The term $\sum_i g_i \mathcal{Q}_m$ is constant and does not influence the disuniformity, wherefore the contribution to the latter is due only to

$$\sum_i g_i \Delta\mathcal{Q}_i , \tag{3}$$

in which the different $\Delta\mathcal{Q}_i$, we repeat, have well-defined signs such that

$$\sum_i \Delta\mathcal{Q}_i = 0 .$$

The expression (3) becomes zero if

a) $g_1 = g_2 = \ldots = g_n$ since, as it has been said, $\sum_i \Delta\mathcal{Q}_i = 0$.

b) Every $\mathcal{Q}_i = \text{const}$ for any point at which scintillation takes place.

In general these two conditions have never been verified; the first because of the unfailing difference in the value of g between one P.M. and another. (It is well to remember that in every case the values of g_i can be adjusted by a suitable variation of the tension in one of the intermediate states of multiplication). Even in the case of a sole P.M. the uncertainty about the value of g is equally dependent on the different sensitivity point by point of the photocathode. The second condition is not verified because light is distributed differently on the photosensitive face with the varying of the scintillation position, whatever may be the counter form and its dimensions.

To estimate (3), we may take as known the g_i mean value g_m with a $p\,\%$ error, and we may imagine as known every single deviation $\Delta\mathcal{Q}_i$, in value and sign, for whatever point of the counter, in respect to the mean value $\mathcal{Q}_m$ of the luminous flux received by each tube.

The expression (3) then becomes

$$g_m \sum_i \Delta\mathcal{Q}_i \pm p \cdot 10^{-2}\, g_m \sum_i |\Delta\mathcal{Q}_i| .$$

If the counter is at uniform collection, with $\sum_i \mathcal{Q}_i = 0$, the contribution to disuniformity is reduced to

$$p \cdot 10^{-2}\, g_m \sum_i |\Delta\mathcal{Q}_i| , \tag{4}$$

where the value of $\Delta\mathcal{T}_i$ depends on the point at which scintillation takes place. The meaning of (4) is clear; it represents the maximum possible, positive or negative, separation from its mean value $\sum_i g_m\mathcal{T}_m$ of the response.

The maximum percentage value of disuniformity obtained is then

$$\frac{p\sum_i|\Delta\mathcal{T}_i|}{\mathcal{T}}. \tag{5}$$ (*)

It can be seen that, case by case, it is necessary to know the point-by-point value of $\Delta\mathcal{T}_i$ in order to estimate what uniformity can be expected. This means, in other words, that we have to find the light distribution function on the photosensitive walls.

The curves of Fig. 4 refer to a counter of the type which has been discussed previously, and represented schematically in Fig. 5. The pipe light of such a counter, even with a length $l = d\cdot\mathrm{tg}\,\gamma$, shows two parallel, plane-surfaced cuts of a length $h < l$. With this arrangement, the uniformity of light collection has again been realized, but its distribution on the photosensitive surface takes place as if this were found on the upper plane which limits the cuts, and is therefore greatly altered in respect to the pipe light without cuts. The cuts are in practice necessary for the connexion between the collection surface and that of the P.M. We have been referring to a counter of this type, because it corresponds to the one we actually used and which will be discussed later on. Fig. 4 reports in ordinates, in arbitrary unit, the quantity of light received by each tube and in abscissae the distance $Y-h$, where it is $Y=y+l$. The curves have been calculated (see Appendix II), for the points of the straight line drawn which corresponds to a P.M. axis.

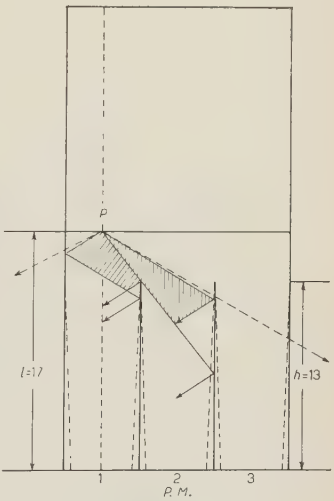


Fig. 5.

From (5) and the curves of Fig. 4 and admitting that $p=10$ and $Y-h=4$ cm, a maximum disuniformity of $\simeq 8\%$ is to be expected, as a consequence of

(*) This expression always leads to an overestimation of disuniformity due to this cause, while the contribution to disuniformity of all the P.M.'s is supposed as summed. This condition in practice can be avoided with a convenient choice of P.M.'s, wherefore we are permitted to think that the values furnished by (5) are increased in respect to the real ones.

light distribution on the photosensitive surface. The family of curves in Fig. 6, where p is the parameter, gives the maximum disuniformity in function of $(l-h)/(c/3)$, where l is the value of Y for which there is maximum disuniformity, and the $l-h$ differences has been normalized at $c/3$. The arrow is to indicate the abscissae whereby the maximum disuniformity in the counter we constructed is obtained.

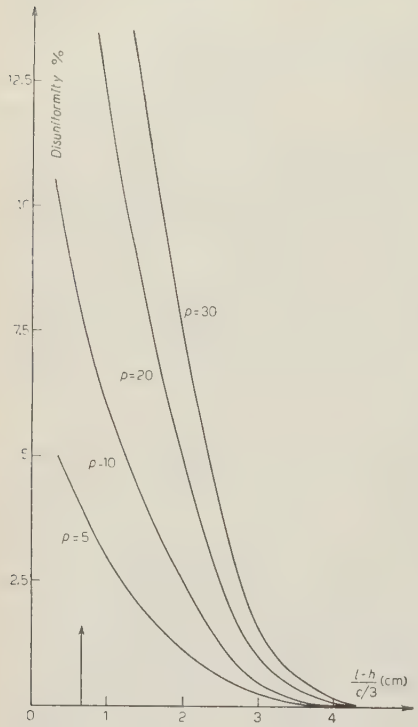


Fig. 6.

If the photosensitive surface is that corresponding to the photocathode of a sole P.M., this estimation can still be important ([9,10]). In this case it is sufficient to think of subdividing the photocathode surface in a convenient number of equal areolae, sensitivity being equal in each. Repeating the preceding arguments, we find that the maximum relative disuniformity is still given by (5) where the $\sum$ is extended to the areola number in which the cathode surface is subdivided, and the value of p expresses the uncertainty of the information of g on every areola considered (*). Estimations of the influence of p on the resolutive power in amplitude of a crystal-spectrometer carried out with a sole P.M. 5819 have been made by Borkowski and Clark ([11]) and Godlove and Wadey ([12]). The results of these last authors indicate that photocathode surface disuniformity is not determinant to the resolutive power effects. This essentially depends on the average sensivity of the photocathode. To state the matter more exactly, the resolutive power depends fundamentally on the

([10]) G. F. J. Garlick and G. Y. Wright: *Proc. Phys. Soc.*, B **65**, 415 (1952).

(*) Referring once more the already-mentioned notations of Birks, there is a contribution to the uncertainty of g on every areola by the product

$$C_{PE} f(v_P) c'$$

where c' is the collection efficiency factor of the photoelectrons from the cathode at the first dynode of the multiplication system.

([11]) C. J. Borkowski and R. L. Clark: *Rev. Scient. Intsr.*, **24**, 1046 (1953).

([12]) T. F. Godlove and W. G. Wadey: *Rev. Scient. Intsr.*, **25**, 1 (1954).

statistical fluctuations of the photoelectrons, which are smaller as the mean sensivity is greater.

For a counter which is not optically uniform the expression of the maximum relative disuniformity becomes

$$\frac{\Delta \mathcal{N}}{\mathcal{N}_m} + p \sum_i \frac{\Delta \mathcal{N}_i}{\mathcal{N}_m},$$

where the first term is due to disuniformity in collection (depending on the shape of the counter and on absorption), and the second to light distribution disuniformity.

4. – Uniformity of a Counter with Reflecting Walls.

A counter with reflecting walls and suitably arranged collection surface (arranged so as to receive sooner or later, after successive reflections, all the light emitted in scintillation) would have a uniform collection if it were not for scintillator absorption and the wall reflection coefficient not being equal to one. In fact ([4]), the reflection coefficient is always less than one, and, moreover, is selective for the different wavelengths. The rays which leave the scintillator with a sufficiently large φ angle reach the collection surface after many reflections, and undergo an attenuation through reflection and absorption which grows with increasing distance of the scintillation from the collection surface. It is evident then that such a counter must be very disuniform, especially because of the rays escaping with a large φ. In fact there will be a great difference between the light collected, in the case in which scintillation takes place near the collection surface, and that in which scintillation takes place far off.

A substantial improvement can be obtained by eliminating those rays which are greatly inclined, since they are the preponderant cause of disuniformity. This object can be reached by adopting the parallelepiped shape with at least one entire collection face, such a face to be united to the photosensitive one by means of a *transparent* lateral-surfaced light pipe, not inferior in length to $d \operatorname{tg} \gamma$, and similar to that examined in the previous case for transparent-walled counters. The influence of the pipe light is that of reducing the light which is incident on the photosensitive surface to that corresponding to the case of the transparent-walled counter.

For reasons which have been explained previously, the use of such a pipe light excludes the consideration of any shape different from the parallelepiped one. In order to carry out a comparison between two parallelepiped-shaped counters, one with reflecting walls and the other with transparent ones, having

equal dimensions and employing the same scintillator, it is sufficient to take into consideration the exponent A of (2).

In the second case let us remember that it is

$$A_t = y\{\alpha_s(\lambda) + C_t\} + l\alpha_p(\lambda), \tag{6}$$

where it is

$$C_t = 2\{\alpha_w(\lambda) - \alpha_s(\lambda)\}\left(\frac{s_1}{b} + \frac{s_2}{c}\right).$$

In the first case it is easy to show that it is

$$A_r = y\{\alpha_s(\lambda) + C_r\} + l\alpha_p(\lambda), \tag{7}$$

where it is

$$C_r = \left(\frac{\sin\varphi}{b} + \frac{\sin\theta\cos\varphi}{c}\right)\alpha_r(\lambda, \varphi, \theta)$$

and α_r represents the absorption coefficient due to reflection, related to the reflection coefficient k by

$$k(\lambda, \varphi, \theta) = \exp\left[-\alpha_r(\lambda, \varphi, \theta)\right].$$

We may state that α_r, being in every case $k < 1$, will always result positive and, then, we shall always have $C_r \geqslant 0$.

From a comparison of (6) and (7), it can be seen that two different situations can be obtained:

a) $\alpha_s(\lambda) \geqslant \alpha_w(\lambda)$. That corresponds to $C_t \leqslant 0$. It will then always result $A_t \leqslant A_r$;

b) $\alpha_s(\lambda) < \alpha_w(\lambda)$. It will then result $C_t > 0$. In this case a comparison between the two types of counters is not immediate.

From this examination it can be affirmed that the counter with transparent walls is, in situation *a*), always the most convenient of those which have been examined up to the present.

This was the case, for example, of the counter which we constructed, in which the walls were of plexiglass, and the scintillator of phenylcyclohexane + + terphenyl + diphenylhexatriene.

The use of glass walls (which is necessary when the scintillator damages the plexiglass) causes us to fall, in many cases, into situation *b*), for which reason the convenience of the use of reflecting walls cannot be here rejected.

We considered that a strict calculation of the uniformity of counters with

reflecting walls would not be necessary. We think in fact that, for counters constructed with the best reflecting surfaces, those which were completed for transparent walls will be still sufficiently approximate.

5. – Uniformity of a Counter with Diffusing Walls.

In order that a counter with diffusing walls have uniform collection it is necessary (leaving aside for a moment the question of scintillator absorption) for it to have an internal spherical shape and a collection surface which is small in respect to the spherical one, in such a way as to realize the conditions of Ulbricht's sphere. Strictly speaking, such conditions can never be realized. In fact in such a counter

a) the collection surface is never negligible in respect to the spherical one (unless there is *a priori* renunciation of a good collection yield);

b) the collection surface always collects a certain percentage of light which comes directly from scintillation. When scintillation takes place very close to the collection surface, the direct light amounts, practically speaking, to the half of that emitted. At greater distances, on the contrary, the direct light becomes negligible.

If, moreover, absorption is taken into account, a further very grave cause of disuniformity is introduced. It is easy to see that the said absorption can be taken into account by conveniently diminishing the diffusion factor μ of the spherical walls to a value of $\mu' < \mu$ (*).

An evaluation of the disuniformity of this counter can be made if we consider as still valid the results of Ulbricht's ideal sphere relatively to the light which is not directly collected by the collection surface. Let us indicate with (Fig. 7):

— R the radius of the sphere;
— r the collection surface radius (supposed as circular);
— y the distance of scintillation P from the collection surface.

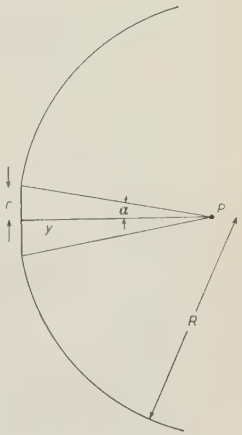


Fig. 7.

(*) The new diffusion coefficient can be considered as expressed by

$$\mu' = \mu \exp\left[-\alpha_s R_m\right]$$

where μ is the wall diffusion coefficient;
α_s is the scintillator absorption coefficient;
$R_m = (4/3)R$ the mean course of the light between two diffusions within the sphere of radius R.

The percentage of direct light (overestimated since we suppose that all the rays originating from P follow the same range in order to arrive on the collection surface) will be

$$R_t = \frac{2\pi(1-\cos\alpha)}{4\pi}\exp[-\alpha_s y]10^2 = \frac{1}{2}\left[1-\frac{y}{\sqrt{y^2+r^2}}\right]\exp[-\alpha_s]10^2\,,$$

where, given the indicative character of this estimation, α_s represents a convenient mean value of the scintillator absorption coefficient relative to the different wavelengths of the emission spectrum. The percentage of diffused light which will reach the collection surface by diffusion will be

$$R_d = \frac{4\pi - 2\pi(1-\cos\alpha)}{4\pi}\,\frac{\pi r^2/4\pi R^2}{(1-\mu')+(\pi r^2/4\pi R^2)}\,10^2 =$$

$$= \left(1+\frac{y}{\sqrt{y^2+r^2}}\right)\frac{r^2}{8R^2(1-\mu')+r^2}\,10^2.$$

Neglecting r^2 in respect to R^2, since $r^2 \ll R^2$, the percentage of light collected will be:

$$(8)\qquad R_t + R_d =$$

$$= \left\{\frac{1}{2}\left(1-\frac{y}{\sqrt{y^2+r^2}}\right)\exp[-\alpha_s y] + \frac{r^2}{8R^2(1+\mu')}\left(1+\frac{y}{\sqrt{y^2+r^2}}\right)\right\}10^2\,.$$

Making a numerical case, by assuming the values:

$$r = 2\ \text{cm}\,, \qquad R = 10\ \text{cm}\,, \qquad \mu = 0.98\,, \qquad \alpha_s = 0.08\ \text{cm}^{-1}$$

we obtain respectively for the two limit positions $y = 0$ and $y = 2R$

$$(R_t + R_d)_{y=0} \simeq (50 + 0.5)\ \%$$

$$(R_t + R_d)_{y=2R} \simeq 1\ \%\,.$$

Such a counter then shows itself to be completely disuniform. It may be of interest to observe that when the condition $y^2 \gg r^2$ is satisfied, (8) becomes

$$(9)\qquad R_t + R_d \simeq \frac{r^2}{4R^2(1-\mu')}\,100\ \%\,,$$

which does not depend on the position at which scintillation takes place.

This situation can be realized practically by placing a suitably-shaped scintillator in a convenient region of the sphere in respect to the collection surface, provided that the remaining region of the sphere be filled with a well-transparent medium having the same refractive index as the scintillator. In this case, if the absorption coefficient of this substance is less than α_s, μ' will increase, and so improve the yield.

Should it be desired to examine shapes other than the spherical one by using diffusing walls, the situation will worsen still further, since the symmetry which exists in the case of the diffusing sphere is lacking. In this connexion M. MANDÒ's [13] conclusions can be considered valid, according to which the results of Ulbricht's sphere can be extended approximatively to other shapes, provided that these are sufficiently regular and that the three maximum dimensions therein are not very different.

6. – Experimental Proofs.

We constructed a parallelepiped-shaped liquid counter, measuring 18 cm × 18 cm × 2 cm, with plexiglass transparent walls having a thickness $s=2.0$ mm, in which light collection was carried out on an entire wall (one of those

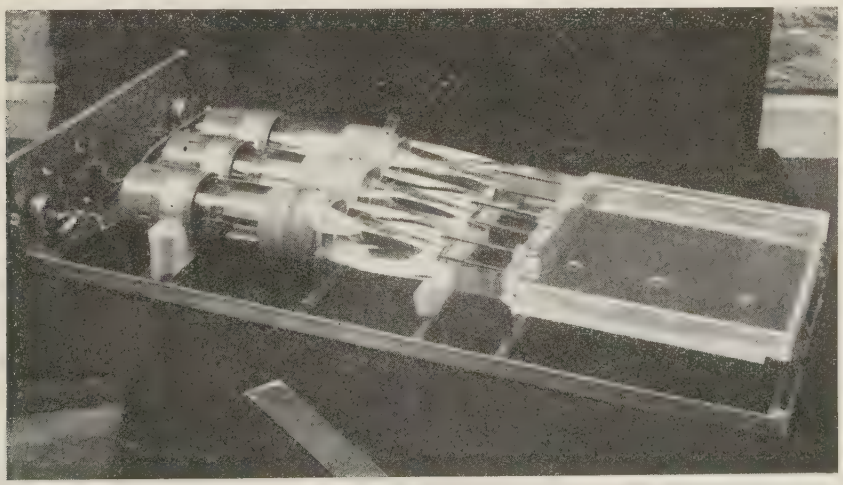


Fig. 8.

of the smaller area). To this a pipe light (Fig. 8), also of plexiglass, with a length of $l=17$ cm, was applied. For the joining of the rectangular collection face with the photosensitive surface constituted by 3 P.M. 5819 photocathodes,

[13] M. MANDÒ: *Nuovo Cimento*, **12**, 5 (1954).

the pipe light was suitably cut with two long incisions $h = 13$ cm. As scintillator we used a solution of 3 g of terphenyl + 1 litre of phenylcyclohexane + 10 mg of diphenylhexatriene.

On the basis of the considerations discussed above, the disuniformity estimated for the effect of absorption shows itself to be $\simeq \pm 11\%$ in respect to the mean value. So, taking the preceding calculations as a basis we get a maximum yield of $\sim 16\%$ for scintillations near the pipe light and a minimum yield of $\sim 13\%$ for those which take place at the maximum distance therefrom. In calculating disuniformity due to absorption we supposed the emission spectrum to be of constant intensity on all wavelengths and to be limited between the values $\lambda = 3300$Å and $\lambda = 4700$Å. The solution and plexiglass absorption coefficients were measured directly with a Bekmann's spectrometre on samples with a thickness of 10 cm. The results of these measurements are reported in Fig. 9.

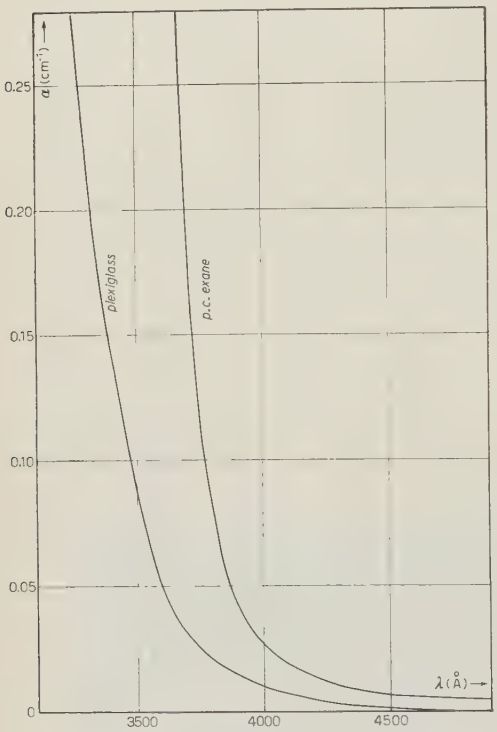


Fig. 9.

Analogous calculations, carried out for the same counter furnished with an aluminium reflecting plate (mean reflection coefficient 0.95) which was applied to the face opposite to the collecting one, permitted an estimation of disuniformity due to absorption as $\simeq 7\%$. The maximum yield in this case was estimated at $\simeq 20\%$ and the minimum at $\simeq 17\%$. The curves in Fig. 10 report the behaviour, in arbitrary units, of the counter's response as a function of the distance of the scintillation. They were calculated, for the two cases considered, by taking absorption only into account.

To the disuniformity due to absorption must be added that due to light distribution on the photosensitive surface. The comparison among our P.M. values of g allowed us to estimate a value of ~ 10 for p. From the curves of Fig. 4 we calculated a maximum disuniformity of $\sim 8\%$ in the counter without a reflecting plate.

The experimental proofs were carried out with a device which is schematized in Fig. 11. A ^{60}Co source of 50 millicurie was collimated with a block

of Pb of the thickness of 20 cm, in which a hole with a diameter of 8 mm was bored. The counter was tested in several points by the collimated small beam and the direct current escaping from the P.M.'s measured with a galvanometer. The currents of each phototube were sufficiently weak as to consider that eventual saturation space charge phenomena would not take place. For these reasons we consider that the results obtained are still valid in an impulsive operation.

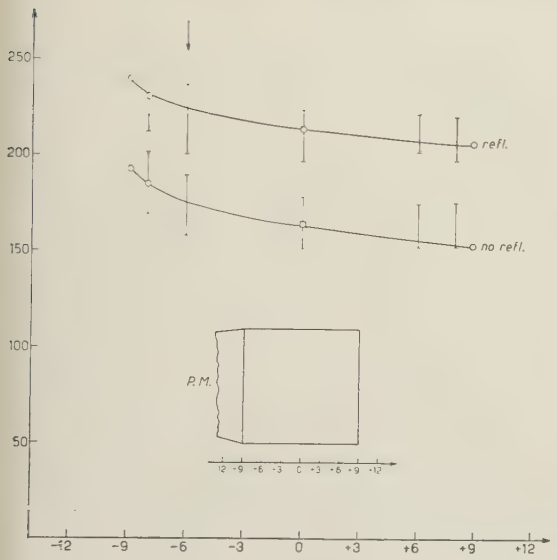


Fig. 10.

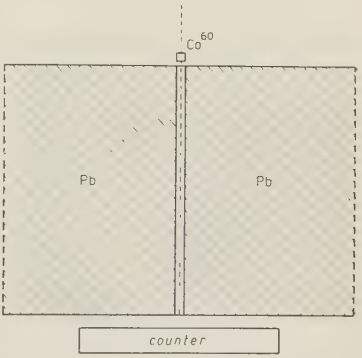


Fig. 11.

Equal response curves, relative to the two cases treated above, are shown in Fig. 12. We limited the estimation of the equal response curves to a surface of the counter of 16×16 cm, leaving out a border of 1 cm because, on

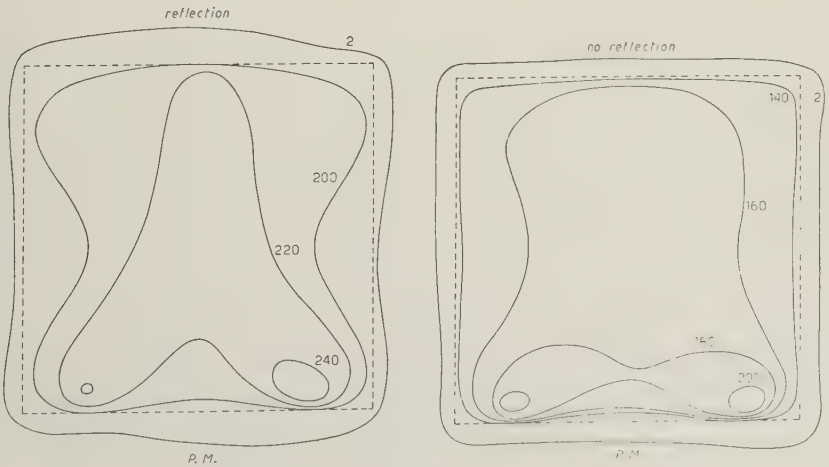


Fig. 12.

this edge, the influence of the beam width, which in our case was estimated at ~ 2 cm, made itself felt.

Fig. 10 also reports the maximum and minimum values of the response measured along axes parallel to the collection face and normalized in respect to the point signed by a cross. The response variations along these axes are ascribable to the sole disuniformity of light distribution. The results of the measurements, relative to the maximum variation of the response, allow an estimation of this disuniformity in the values $\sim 5\%$ and $\sim 4\%$ respectively for the counter without and with a reflecting plate.

The experimental results relative to the counter without a reflecting plate show a disuniformity of $\sim 14\%$. This value is certainly in default, because the responses on the counter edge near the collection face were not taken into consideration. The disuniformity calculated appears, on the contrary, expressed by the sum of the disuniformity due to absorption ($\sim 11\%$) and that due to light distribution ($\sim 8\%$). If however, we take into account the fact that the latter contribution was largely overestimated and can be reduced by some units, we see that the experimental results are in good agreement within the limits of the approximations made, with the theoretical ones.

In the case of the counter with a reflecting plate, we found a disuniformity of $\sim 11\%$ which was still in good agreement with our previsions. Finally we should like to observe that the yield increase due to the addition of the reflecting plate is also in good agreement with the increase calculated.

7. – Light Collection in Time.

It is important to make some observations on the temporal development of light collection in an impulsive operation in order to underline the advantages which, even in this connexion, a counter with transparent walls presents in respect to those furnished with other types of walls. In a parallelepiped-shaped counter the light collected is due to two contributions; the first relative to the light which is propagated towards the collection face, and the second which goes in the opposite direction and reaches the collection surface after having undergone reflection on the opposite surface parallel to it. This second contribution superimposes itself on the first, shifted in respect thereto by a time dependent on the scintillation position and counter dimensions. A strict calculation of the temporal function of collection in order to estimate the time T necessary for the collected light to reach its maximum value (practically 99%) turns out to be very laborious, even if rather simple from a theoretical point of view.

Appendix III reports the calculation of the collection function in a particular case. The results of this calculation, nevertheless, allow us in general to establish an inferior limit for the collection time, a limit which can be expressed by

$$T = \eta\tau + t_1 = \eta\tau + \frac{2a}{v} \tag{10}$$

where τ is the decay time of the phosphor, t_1 the maximum shift time between the two contributions, v is the propagation speed of light in the medium and η is a convenient number depending on the counter dimensions, which in practice goes from 5 to 10. The value of η, all other conditions being equal, is always less in a counter with transparent walls than in one with reflecting or diffusing walls. In fact, in these last cases, a contribution is made to the pulse by luminous rays which travel a long way before reaching the photosensitive surface, rays which are for the most part eliminated in the case of transparent walls. The tension pulse at the P.M. anode has the same proceeding as light collection if the time constant of the output integral circuit is much greater (theoretically infinite) than T. This condition must be realized when a proportional counter is required.

We propose to assume the time $2T$ as the resolutive time value for such a counter; T also suggests the pulse formation time value.

We wish finally to stress that the estimation of t_1 is very important whenever it is necessary to use large counters.

8. – Conclusion.

From what has been said in the preceding paragraphs, we can draw some conclusions, which are generally valid, and which refer to the construction of minimum-disuniformity scintillation counters, the dimensions of which are some decimetres.

a) The parallelepiped shape, with transparent walls, and at least one entire collection surface connected to the photosensitive surface with a suitably long pipe light is convenient. With two entire photosensitive faces, which are opposite one to the other, uniformity as well as yield is improved. An intermediate situation is obtained by substituting a reflecting plate for a photosensitive face.

b) All other conditions being equal, a counter with reflecting walls is always more disuniform than one with transparent walls. A considerable improvement in uniformity can be obtained in this case by using a pipe light with transparent walls. Such a counter gives a performance which is not far off that of the one with transparent walls. It can prove useful when, for technical reasons, transparent walls cannot be realized.

c) A counter with diffusing walls can be uniform only if the conditions of Ulbricht's sphere are obtained, and if the scintillation zone is opportunely limited. This must never be too near the photosensitive surface. In this case the counter yield is very low and is inferior to the one obtained with the preceding arrangements.

d) In the best of the solutions proposed, absorption and non-regular light distribution on the photosensitive surface always remain present as causes of disuniformity. These play roles of different importance according to the dimensions of the counters. For counters which measure some decimetres the first cause prevails over the second.

* * *

Our grateful tkanks are due to Mr. A. COLLINA for his construction of the counter described in the present paper.

APPENDIX I.

We shall carry out the integration of (2) in relation to the two contributions ΔR_t and ΔR_b, subdividing the integration field in a suitable manner. We can at once see from Fig. 13 that

$$\text{(A.1)}\quad \Delta R_t = \frac{1}{\pi}\left[\int_{\varphi=0}^{\pi/2}\int_{\theta=0}^{\pi/2} f(A,\theta,\varphi)\,\mathrm{d}\varphi\,\mathrm{d}\theta - \int_{\varphi=\pi/2-\gamma}^{\pi/2}\int_{\theta=0}^{\pi/2} f(A,\theta,\varphi)\,\mathrm{d}\varphi\,\mathrm{d}\theta - \int_{\varphi=0}^{\gamma}\int_{\theta=\theta'_\varphi}^{\pi/2} f(A,\theta,\varphi)\right.$$

for $A(y,\lambda)$ and

$$\text{(A.2)}\quad \Delta R_b = \Delta R_t - \frac{1}{\pi}\int_{\varphi=0}^{\gamma}\int_{\theta=0}^{\theta''_\varphi} f(A,\theta,\varphi)\,\mathrm{d}\varphi\,\mathrm{d}\theta$$

for $A(2a-y,\lambda)$, where it is

$$f(A,\theta,\varphi) = \exp\left[-A/\cos\varphi\cos\theta\right]\cos\varphi$$

and θ'_φ and θ''_φ are such as to satisfy the relations

$$\cos\theta'_\varphi = \frac{\sqrt{\sin^2\gamma - \sin^2\varphi}}{\cos\varphi};$$

$$\sin\theta''_\varphi = \frac{\sqrt{\sin^2\gamma - \sin^2\varphi}}{\cos\varphi}.$$

Given the difficulty of such integrations, these have been performed by the graphic method.

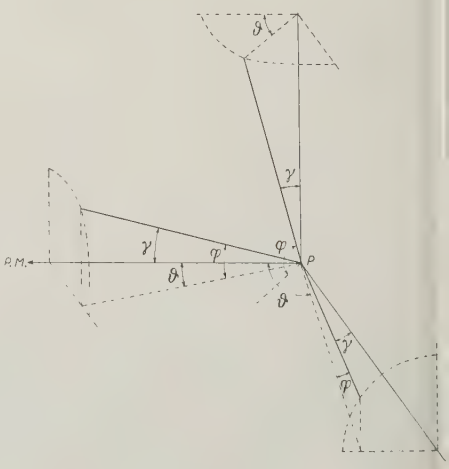


Fig. 13.

Appendix II.

Let a trirectangular parallelepiped (Fig. 14) made of transparent material be given; at a point P therein is placed a pointform source which emits light in all directions. We want to calculate the distribution of the luminous flux which falls on the face situated on the plane π. It is a question funda-

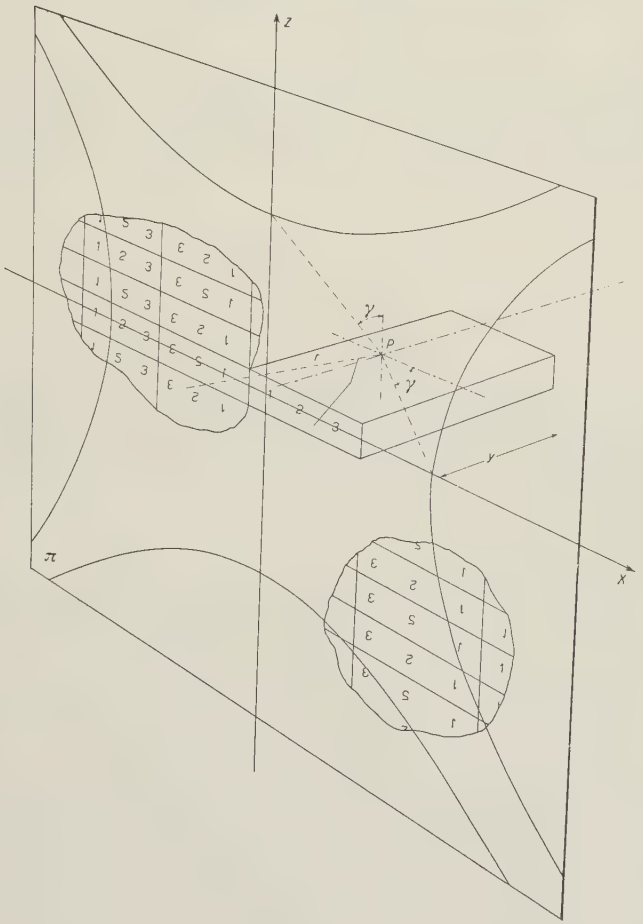


Fig. 14.

mentally of estimating the fluxes collected by each of the square surfaces, in which the counter collection face can be considered divided, each face corresponding to the photo-cathode of a P.M.. (In the case of the figure the P.M. number is three).

The problem can be resolved by dividing the plane π into many contiguous

squares, equal to those in which the collection surface was divided. When a ray escapes from P in a certain direction, it will, after successive reflections, fall on one of the squares of the collection face. If the ray is imagined as propagating in a direct line without undergoing reflections, it will fall on the plane π within a square corresponding to that of the collection surface. The ray r in Fig. 14 illustrates this situation.

The quantity of light which invests the element of the area $\mathrm{d}x\,\mathrm{d}z$ on the plane π will then be given by

$$\mathrm{d}\mathcal{Q} = \frac{\mathcal{Q}}{4\pi} \frac{Y\,\mathrm{d}x\,\mathrm{d}z}{(x^2 + Y^2 + z^2)^{\frac{3}{2}}}, \tag{A.3}$$

where $\mathcal{Q}$ is the total quantity of light emitted by the luminous source, Y the distance of this from π and x and z the coordinates of the surface element. The total light collected by one of the squares in which the photosensitive surface was divided is given by:

$$\mathcal{Q}_i = \frac{\mathcal{Q}}{4\pi} \sum_{1}^{n} {}_j \int_{z_{1j}}^{z_{2j}} \int_{x_{1j}}^{x_{2j}} \frac{Y\,\mathrm{d}x\,\mathrm{d}z}{(x^2 + Y^2 + z^2)^{\frac{3}{2}}}, \tag{A.4}$$

where x_{1j}, x_{2j}, z_{1j}, and z_{2j} are the coordinates of the vertices of one of the generic squares of the plane π corresponding to the one designated on the collection surface. The four branches of the hyperbola, which represent the intersections of the four cones escaping from P with the axes parallel to the plane π and which have a semi-aperture γ, have the equations:

$$x^2 \operatorname{tg}^2 \gamma = Y^2 + z^2 \qquad \text{and} \qquad z^2 \operatorname{tg}^2 \gamma = Y^2 + x^2$$

in the reference system chosen. They represent the variation limits of the coordinates x_j and z_j, in the sense that in the summing-up (A.4) there must be computed only those squares (or portions thereof) which lie in the region of the plane π between them. For every one of the areas in which the collection surface has been divided and is in agreement with every value of Y chosen on an axis normal to π, equ. (A.4) can be graphically estimated in sufficiently rapid manner and with good approximation.

Since disuniformity in distribution is due mainly to the light which reaches the collection face almost directly, i.e. after few reflections, the graphic computation can be arrested for an area round the centre of the reference system which is not too extended, while considering the rest of the light collected by the face to be uniformly distributed. Fig. 4 shows the results obtained for the parallelepiped-shaped counter with transparent walls described above.

Appendix III.

Our present purpose is that of calculating the collection temporal function of photons for a scintillation which takes place at a distance Y, by the photosensitive face. We shall refer to a counter with reflecting walls, parallelepiped in shape with one entire photosensitive face, and we shall suppose there to be no absorption. The considering of absorption does not meaningly complicate the calculations, but only makes them more laborious. On the other hand these hypotheses allow us equally to make an estimation of the inferior limit of the T value of (10); and this is the estimation which interests us.

Let us suppose (Fig. 15) that a scintillation takes place in P. Assuming as the time origin (all measured in τ units) the time in which the first photon reaches 0, the number of photons which arrive inside the solid elementary angle $d\omega$, the photosensitive surface up to the time t is given by:

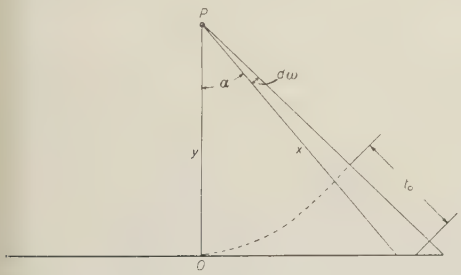


Fig. 15.

$$d\mathcal{N} = \frac{\mathcal{N}}{4\pi}\left[1 - \exp\left[-(t - t_0)\right]\right] d\omega \,,$$

where t_0 is the time taken in going over the distance $x - Y$. The solid elementary angle $d\omega$ can be expressed by

$$d\omega = \frac{2\pi Y \tau v \, dt_0}{(Y + \tau v \cdot t_0)^2} \,,$$

where v is the light propagation velocity in the medium considered. The total number of photons collected by the photosensitive surface up to the time t, will be:

$$\mathcal{N}(t) = \frac{\mathcal{N}}{4\pi}\int_0^t \left[1 - \exp\left[-(t - t_0)\right]\right] \frac{2\pi Y \tau v}{(Y + \tau v t_0)^2} dt_0 \,. \tag{A.5}$$

The superior limit t of t_0 is established by the fact that it must always be $t_0 < t$.

The integration of (A.5) leads to

$$\mathcal{N}(t) = \frac{\mathcal{N}}{2}\left\{1 - e^{-t} - \Lambda e^{-\Lambda t}\left[E_i(\Lambda + t) - E_i(\Lambda)\right]\right\} \tag{A.6}$$

where $\Lambda = Y/\tau v$.

As can be seen from (A.6) and has already been said before, the collection function depends on Y, the distance of scintillation from the photosensitive surface.

The (A.6) refers to a sole semispace of emission. The contribution due to the other semispace is calculated with a relation analogous to (A.6), where,

in the place of Y, must be put $2a - Y$, and where of course the times beginning with the same origin must be counted.

For purposes of demonstration, we have reported in Fig. 16 some behaviours of (A.6) for various values of Λ. As can be seen the time necessary for reaching the maximum depends in a marked degree on Λ.

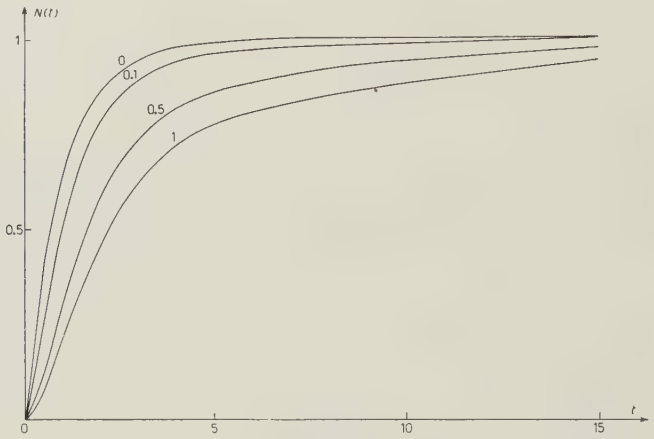


Fig. 16.

In the case of transparent walls, collection appears without doubt more rapid, since the tail contributions make themselves less felt. When we take the absorption into account, it can be seen that this cuts off the contributions due to the longer paths and so further reduces the slow part of the signal.

We consider therefore that only in particular conditions the time T does reach the total value $\approx 10\tau + \tau_1$.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

Über den Einfluß von Fehlern des magnetischen Feldes auf die Betatronschwingungen im Synchrotron mit starker Stabilisierung (1).

G. Lüders

Max-Planck-Institut für Physik - Göttingen
Europäische Organisation für Kernphysikalische Forschung (CERN)

(ricevuto il 25 Marzo 1955)

Inhalt: Einleitung. *a*) Das Prinzip starker Stabilisierung *b*) Typen von Feldfehlern und ihre Auswirkungen. – I. Allgemeine Behandlung. 1. Fehler im magnetischen Führungsfeld. *a*) Gleichung der geänderten Gleichgewichtsbahn. *b*) Spezielle Fälle. *c*) Anwendung auf die «synchrone Bahn». 2. Fehler im Feldindex. *a*) Störungsrechnung für Fehler im Feldindex. *b*) Gleichungen für den instabilen Bereich. *c*) Flanken der instabilen Bereiche. 3. Verschiedene Probleme. *a*) Gleichzeitige Wirkung von Fehlern im Führungsfeld und im Feldindex. *b*) Ähnlichkeitsgesetze. 4. Zusammenfassung von Teil I. – II. Statistische Auswertung. 1. Mittelwerte. *a*) Geänderte Gleichgewichtsbahn, Vorbemerkungen. *b*) Geänderte Gleichgewichtsbahn, Vernachlässigung der Amplitudenschwankung entlang des Umfangs. *c*) Geänderte Gleichgewichtsbahn, Amplitudenschwankung entlang des Umfangs. *d*) Instabile Bereiche. 2. Wahrscheinlichkeitsverteilungen. *a*) Geänderte Gleichgewichtsbahn. *b*) Instabile Bereiche. 3. Zusammenfassung von Teil II. – Anhang 1: Bewegungsgleichungen in linearer Näherung. *a*) Vernachlässigung der Beschleunigung. *b*) Berücksichtigung der Beschleunigung. – Anhang 2: Zur störungsmässigen Behandlung kleiner Änderungen des Feldindex. *a*) Das Innere eines stabilen Gebiets. *b*) Die Grenze eines stabilen Gebiets. *c*) Instabile Bereiche infolge Fehlern im Feldindex.

(1) Diese Arbeit gibt im wesentlichen eine Zusammenfassung von Forschungsberichten, die vom Verfasser für die Europäische Organisation für Kernphysikalische Forschung (European Organisation for Nuclear Research, CERN) bis zum Herbst 1953 angefertigt wurden. Die Untersuchungen wurden durchgeführt am Institut für Theoretische Physik der Universität Kopenhagen in hauptamtlicher Tätigkeit für die Theoretische Studiengruppe von CERN und am Max-Planck-

Einleitung.

a. *Das Prinzip starker Stabilisierung.* – Das Synchrotron (2) stellt einen Beschleuniger dar, bei dem die geladenen Teilchen, z.B. Protonen, durch ein magnetisches Führungsfeld in der Nähe einer in sich zurücklaufenden Bahn herumgeführt werden. Die Teilchen durchlaufen elektrische Wechselfelder und nehmen bei richtiger Phasenlage in Bezug auf diese Wechselfelder während eines Beschleunigungszyklus fortgesetzt Energie auf. Eine bestimmte Teilchenbahn im Synchrotron wird als Gleichgewichtsbahn bezeichnet; sie stellt eine kreisähnliche geschlossene Kurve dar und verläuft bei den üblichen Entwürfen in der horizontalen Ebene. Das magnetische Führungsfeld am Ort der Gleichgewichtsbahn ist dann senkrecht gerichtet; diese Richtung sei im folgenden als z-Richtung bezeichnet (von der Möglichkeit einer Feldkomponente parallel zur Gleichgewichtsbahn sei abgesehen). Damit die Gleichgewichtsbahn für ein Teilchen der Ladung e und mit dem Impuls p eine mögliche Teilchenbahn darstellt, muß das Magnetfeld an einer bestimmten Stelle der Gleichgewichtsbahn mit dem Krümmungsradius ϱ an dieser Stelle in der Beziehung stehen

$$H_z = -\frac{pc}{e\varrho} \tag{E.1}$$

(c = Lichtgeschwindigkeit).

Infolge der Beschleunigung nehmen Geschwindigkeit und Impuls der Teilchen zu. Die Frequenz der Beschleunigungsspannung muß umgekehrt proportional zur Umlaufzeit gewählt werden; während des Beschleunigungszyklus muß sie daher monoton zunehmen. Damit ein auf der Gleichgewichtsbahn startendes Teilchen während des ganzen Beschleunigungszyklus auf der Gleichgewichtsbahn bleibt, muß das Magnetfeld gemäß Gl. (E.1) dem jeweiligen Impuls angepaßt sein und seinerseits monoton zunehmen.

Unter idealen Betriebsbedingungen bleibt ein Teilchen, das mit dem richtigen Anfangsimpuls in richtiger Phasenlage zur Beschleunigungsspannung auf der Gleichgewichtsbahn startet, während des ganzen Beschleunigungszyklus auf der Gleichgewichtsbahn. Dieses Teilchen sei als Gleichgewichtsteilchen, seine momentane Energie als Gleichgewichtsenergie bezeichnet. Für den erfolgreichen Betrieb des Synchrotrons kommt es darauf an, daß auch Teilchen, die mit etwas anderen Anfangsbedingungen (Impuls, Phase und geometrische

---

Institut für Physik, Göttingen, in nebenamtlicher Tätigkeit für die Protonen-Synchrotron-Gruppe. Die Einreichung zur Veröffentlichung wurde durch äußere Umstände verzögert. Kleinere Ergänzungen wurden nach dem Herbst 1953 hinzugefügt; diese wurden vom Verfasser als Mitarbeiter der Protonen-Synchrotron-Gruppe in Genf erarbeitet.

(2) V. Veksler: *Journ. Phys. USSR*, **9**, 153 (1945); E. M. McMillan: *Phys. Rev.*, **68**, 143 (1945).

Anfangsbedingungen) starten, oder später durch Zusammenstöße mit dem Gas und Abweichungen von den idealen Betriebsbedingungen gestört werden, in der Umgebung des Gleichgewichtsteilchens bleiben und ebenfalls den Beschleunigungsprozeß erfolgreich durchlaufen. Es stellt sich daher das Problem der Stabilität. Teilchenbahnen mit abweichenden Anfangsbedingungen müssen oszillatorische Bewegungen um das Gleichgewichtsteilchen herum ausführen. Die Schwingungen von Teilchen richtiger Energie und Phasenlage um die Gleichgewichtsbahn, auf die sich die vorliegende Arbeit im wesentlichen beschränken wird, werden als Betatronschwingungen bezeichnet, da sie im Betatron in ähnlicher Weise auftreten. Teilchen mit einer von der jeweiligen Gleichgewichtsenergie abweichenden Energie und einer abweichenden Phasenlage beschreiben andere Bahnen im Synchrotron; die hiermit verbundene Wegverlängerung oder -verkürzung hat Schwingungen um die Gleichgewichtsenergie und die Sollphase zur Folge. Infolge dieser Phasenstabilität ist es daher nur nötig, die jeweilige Frequenz und das Magnetfeld an einander anzupassen; bei hinreichend großem Scheitelwert der Beschleunigungsspannung stellt sich die stabile Phase dann von selbst ein [3].

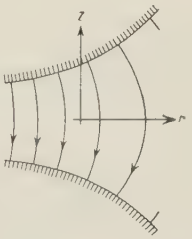


Abb. 1. – Schnitt senkrecht zur Gleichgewichtsbahn mit Angabe der von der Gleichgewichtsbahn aus gemessenen Koordinaten r und z. Polschuhe und magnetische Feldlinien sind eingezeichnet.

Stabilität der Betatronschwingungen, d.h. eine oszillatorische Bewegung der Teilchen um die Gleichgewichtsbahn wird erreicht, indem man das Magnetfeld in geeigneter Weise inhomogen wählt. Bezeichnet man die vertikale Auslenkung eines Teilchens mit z, die horizontale Auslenkung senkrecht zur Gleichgewichtsbahn (fort vom Mittelpunkt des Beschleunigers) mit r (Abb. 1), mißt man ferner die Bogenlänge entlang der Gleichgewichtsbahn in einer Einheit ϑ, die während eines Umlaufs von 0 bis 2π zunimmt, so gelten (bei Vernachlässigung der Beschleunigungsspannung) in linearer Näherung folgende Bewegungsgleichungen

$$\text{(E.2)} \qquad \begin{cases} z''(\vartheta) - n_{zz}(\vartheta) z(\vartheta) - n_{zr}(\vartheta) r(\vartheta) = 0 , \\ r''(\vartheta) - n_{rz}(\vartheta) z(\vartheta) - n_{rr}(\vartheta) r(\vartheta) = 0 , \end{cases}$$

Die Striche bedeuten Ableitungen nach der Größe ϑ. Die Gleichungen sind abgeleitet und im einzelnen erläutert in Anh. 1. Es ist (mit $2\pi R$ als Umfang

[3] Vgl. D. Bohm und L. Foldy: *Phys. Rev.*, **70**, 249 (1946).

der Gleichgewichtsbahn)

$$\text{(E.3)} \qquad \begin{cases} n_{zz}(\vartheta) = \dfrac{R^2}{H_z(\vartheta)\,\varrho(\vartheta)}\,\dfrac{\partial H_z(\vartheta)}{\partial r}\,, \\ n_{rr}(\theta) = -\dfrac{R^2}{H_z(\vartheta)\,\varrho(\vartheta)}\,\dfrac{\partial H_z(\vartheta)}{\partial r} - \left(\dfrac{R}{\varrho(\vartheta)}\right)^2, \\ n_{zr}(\vartheta) = n_{rz}(\vartheta)) = -\dfrac{R^2}{H_z(\vartheta)\,\varrho(\vartheta)}\,\dfrac{\partial H_z(\vartheta)}{\partial z}\,. \end{cases}$$

Daß hier nur Ableitungen der z-Komponente des magnetischen Feldes vorkommen, rührt daher, daß die vertikale und horizontale Komponente des magnetischen Feldes durch die Maxwellschen Gleichungen mit einander verknüpft sind. Der Einfachheit halber soll weiterhin angenommen werden, daß

$$\text{(E.4)} \qquad n_{zr}(\vartheta) = n_{rz}(\vartheta) = 0\,,$$

was z.B. erfüllt ist, wenn das magnetische Feld symmetrisch bezüglich der Ebene der Gleichgewichtsbahn ist. Die Betatronschwingungen in vertikaler und horizontaler Richtung sind dann unabhängig von einander.

Bei dem älteren Stabilisierungsprinzip [4] wird die Gleichgewichtsbahn als Kreisbahn ($\varrho(\vartheta) = R$) und $\partial H_z/\partial r$ längs des Umfangs konstant gewählt. Die Bewegungsgleichungen (E.2) werden dann gewöhnliche Oszillatorgleichungen und man hat oszillatorisches Verhalten der vertikalen und horizontalen Betatronschwingungen, falls gilt

$$\text{(E.5)} \qquad 0 < -\frac{R}{H_z}\,\frac{\partial H_z}{\partial r} < 1\,.$$

Die Wellenlänge der Betatronschwingungen (kurz: Betatronwellenlänge) ist dann für beide Schwingungsrichtungen größer als ein Umlauf. Das neue, von Christophilos, Courant, Livingston und Snyder vorgeschlagene [5] « Prinzip starker Stabilisierung » [6] erlaubt dagegen prinzipiell das Auftreten einer be-

---

[4] D. W. Kerst und R. Serber: *Phys. Rev.*, **60**, 53 (1941).

[5] E. D. Courant, M. S. Livingston und H. S. Snyder: *Phys. Rev.*, **88**, 1190 (1952), im folgenden als CLS bezeichnet. Für die Frage der Priorität vgl. E. D. Courant, M. S. Livingston, H. S. Snyder und J. P. Blewett: *Phys. Rev.*, **91**, 202 (1953).

[6] Verf. möchte sich dem englischen Sprachgebrauch, das Prinzip als « strong focusing principle » zu bezeichnen, nicht anschließen, da dieser Ausdruck (oder die entsprechende deutsche Übersetzung) zu stark an optische Abbildungseigenschaften denken läßt. Der neuerdings benutzte Ausdruck « alternating gradient principle » erlaubt keine einfache deutsche Übersetzung.

liebig großen Zahl von Betatronwellen auf dem Umfang; der Teilchenstrahl wird dadurch stärker zusammengehalten und besitzt bei gleichen Eigenschaften des Injektors einen geringeren Querschnitt (7).

Für die mathematische Formulierung des Prinzips starker Stabilisierung sollen die entkoppelten Bewegungsgleichungen (E.2) (vgl. Gl. (E.4)) vereinfacht geschrieben werden. Es sei nur eine der beiden Schwingungsrichtungen, etwa die horizontale, explizit betrachtet; die Bewegungsgleichung werde geschrieben

$$r''(\vartheta) - n(\vartheta) r(\vartheta) = 0 \,. \tag{E.6}$$

Die Größe $n(\vartheta) \equiv n_{rr}(\vartheta)$ soll als Feldindex bezeichnet werden (8). Das Prinzip starker Stabilisierung beruht dann darauf, daß der Feldindex $n(\vartheta)$ als periodische, sowohl positive wie negative Werte annehmende Funktion längs der Gleichgewichtsbahn gewählt wird

$$n(\vartheta + \vartheta_0) = n(\vartheta) \,. \tag{E.7}$$

Die Periode ϑ_0 muß dabei einen ganzzahligen Bruchteil eines vollen Umlaufs (2π) darstellen

$$\vartheta_0 = \frac{2\pi}{M} \,. \tag{E.8}$$

Die (ganzzahlige) Größe M soll als Zahl der Perioden des Feldindex auf dem Umfang bezeichnet werden. Ist $\varrho(\vartheta)$ periodisch mit derselben Periode, so folgt aus Gl. (E.3), daß für die vertikale Schwingung ein Feldindex mit derselben Periode auftritt, der im wesentlichen das Negative des in Gl. (E.6) vorkommenden ist. Obwohl sich die mathematische Analyse auf die horizontale

---

(7) Bei quantitativen Überlegungen muß beachtet werden, daß — bei Injektion nur während eines Umlaufs — der Injektor nicht etwa allein durch den Öffnungswinkel des in das Synchrotron injizierten Strahls charakterisiert ist. Vielmehr liegt bei dem Zusammenfügen des Vorbeschleunigers und des Synchrotrons ein Problem vor, das dem der stoßfreien Verbindung zweier Wellenleiter sehr ähnlich ist. Unter Benutzung elektrischer oder magnetischer Linsen läßt sich stets eine optimale Anpassung erreichen, bei der die Amplitude der beim Einschießen entstehenden Betatronschwingungen am kleinsten ist.

(8) Diese Definition stimmt mit dem üblichen Sprachgebrauch nicht überein, jedoch schien es sich nicht zu empfehlen, für die in dieser Arbeit auftretende Größe $n(\vartheta)$ einen neuen Namen einzuführen. Die meist als Feldindex bezeichnete Größe ist im wesentlichen mit $n_{zz}(\vartheta)$ (Gl. (E.3)) identisch; vielfach wird noch ein Faktor $(R/\varrho(\vartheta))^2$ abgespalten, was sich für die hier vorzutragende allgemeine Theorie nicht empfiehlt. Bei Anwendung der Ergebnisse dieser Arbeit sollte auf die Unterschiede der Definitionen des Feldindex besonders geachtet werden.

Schwingungsrichtung beschränken wird, muß natürlich beachtet werden, daß Stabilität sowohl der horizontalen wie der vertikalen Schwingungen erforderlich ist.

Hat man irgend zwei Lösungen $r_1(\vartheta)$, $r_2(\vartheta)$ der Differentialgleichung (E.6), so ist die aus ihnen gebildete Wronskideterminante

$$W = r_1(\vartheta) r_2'(\vartheta) - r_2(\vartheta) r_1'(\vartheta) \tag{E.9}$$

unabhängig von ϑ; sie ist von null verschieden, falls die beiden Lösungen linear unabhängig voneinander sind. Ferner ist wohlbekannt, daß sich die Periodizität (E.7) des in der Differentialgleichung als Koeffizienten auftretenden Feldindex $n(\vartheta)$ in einer Art Periodizität der Lösungen wiederspiegelt [9]. Man kann nämlich im allgemeinen zwei spezielle linear unabhängige Lösungen $r_1(\vartheta)$, $r_2(\vartheta)$ (»Floquetsche Lösungen«) finden mit den Eigenschaften

$$r_1(\vartheta + \vartheta_0) = \lambda_1 r_1(\vartheta)\,, \qquad r_2(\vartheta + \vartheta_0) = \lambda_2 r_2(\vartheta)\,. \tag{E.10}$$

Die Faktoren λ_1, λ_2 sind von ϑ unabhängig. Da die Differentialgleichung reell ist, sind beide Faktoren entweder reell oder sie sind zu einander konjugiert komplex. Aus der Konstanz der Wronskideterminante folgt, daß das Produkt $\lambda_1 \lambda_2$ stets gleich eins ist. In der komplexen λ-Ebene liegen die beiden Multiplikatoren also entweder auf der reellen Achse oder auf dem Einheitskreis (Abb. 2). Im ersten Fall hat man für λ_1, $\lambda_2 \neq \pm 1$ neben einer abklingenden Lösung auch eine solche, die während der Umläufe (formal, d.h. innerhalb der linearen Näherung) über alle Grenzen wächst. Eine beliebige Lösung oder — physikalisch gesprochen — eine beliebige Teilchenbahn stellt eine Linearkombination dieser beiden speziellen Lösungen dar. Liegen die Bildpunkte auf der reellen Achse, so gibt es also für λ_1, $\lambda_2 \neq \pm 1$ Bahnen, deren Amplitude (formal) über alle Grenzen wächst; man hat dann Instabilität. Im Falle komplexer, auf dem Einheitskreis liegender Multiplikatoren λ_1, λ_2 bleibt dagegen die Amplitude der Betatronschwingungen auch nach beliebig vielen Umläufen beschränkt; man hat Stabilität. Schreibt man

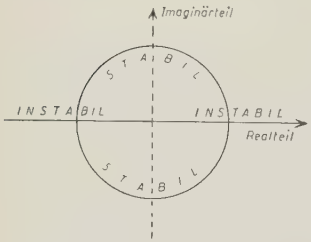


Abb. 2. – Komplexe Ebene der Floquetschen Multiplikatoren λ_1, λ_2; vgl. Gl. (E.10).

$$\lambda_1 = \exp[i\mu]\,, \qquad \lambda_2 = \exp[-i\mu]\,, \tag{E.11}$$

---

[9] Sog. Hillsche Differentialgleichung; vgl. etwa E. KAMKE: *Differentialgleichungen, Lösungen und Lösungsmethoden*, Bd. 1, 410, Leipzig, 1943. Dort weitere Literaturhinweise.

so hat man Stabilität für reelles μ. Macht man die Größe μ eindeutig, etwa durch die Vorschrift $0 \leqslant \mu \leqslant \pi$, so bezeichnet man die Größe $(2\pi\vartheta_0)/\mu$ als (im Winkelmaß gemessene) Betatronwellenlänge. Unter Benutzung von M aus Gl. (E.8) gibt

$$Q = \frac{M\mu}{2\pi} \tag{E.12}$$

die Zahl der Betatronwellenlängen auf einem Umlauf an.

An Hand der Abb. 2 kann man leicht den Einfluß einer stetigen Änderung der Verlaufs der Funktion $n(\vartheta)$ innerhalb eines Periodizitätsintervalls auf die Stabilität diskutieren. Im Anfang liege etwa Stabilität vor. Bei stetiger Änderung des Feldindex werden dann die Bildpunkte auf dem Einheitskreis stetig in einer bestimmten Weise wandern; sie bleiben dabei stets spiegelbildlich zur reellen Achse ([10]). Schließlich werden beide Bildpunkte den Punkt $+1$ oder -1 erreichen. Hier liegt eine Verzweigung der Ortskurve vor, und eine weitere Änderung der Verteilung des Feldindex wird die Bildpunkte im allgemeinen zunächst ein Stück auf der reellen Achse wandern lassen, bis sie umkehren und sich wieder auf dem Einheitskreis bewegen. Man erhält also eine Unterbrechung des stabilen Bereichs durch einen instabilen Bereich.

An den Grenzen des stabilen Bereichs ist entweder $\lambda = +1$ oder $\lambda = -1$. Es gibt dann wenigstens eine Lösung, die Gl. (E.10) befriedigt; sie ist periodisch (für $\lambda = +1$) oder antiperiodisch (für $\lambda = -1$). Falls sich an dieser Stelle auch bei Veränderung des Feldindex durch Multiplikation mit einem konstanten Faktor eine Grenze des stabilen Bereichs befindet, so gibt es nur eine einzige periodische (oder antiperiodische) Lösung

$$r_{\text{per}}(\vartheta+\vartheta_0) = \pm\, r_{\text{per}}(\vartheta) \tag{E.13}$$

und für alle anderen Lösungen gilt

$$r(\vartheta+\vartheta_0) = \pm\,(r(\vartheta) + a r_{\text{per}}(\vartheta)) \tag{E.14}$$

mit einer von der speziellen Lösung abhängigen Konstanten a. Liegt bei der angegebenen Veränderung des Feldindex keine Grenze des stabilen Bereichs vor, so sind alle Lösungen periodisch bzw. antiperiodisch. Der mathematische Beweis dieser Behauptungen kann in Anh. 2*b* gefunden werden.

Ein besonders einfaches Modell für ein Synchrotron mit starker Stabilisierung, das sich wegen der unvermeidlichen magnetischen Streufelder aller-

([10]) Besteht die Änderung des Feldindex in einer Multiplikation mit einem konstanten Faktor, so erfolgt die Wanderung der Bildpunkte monoton mit diesem Faktor, also ohne Umkehr der Richtung, wie in Anhang 2*a* gezeigt werden wird.

dings nicht streng verwirklichen läßt, erhält man, wenn man jede Periode des Feldindex in zwei gleichlange Stücke mit einem Feldindex gleichen Betrages aber entgegengesetzten Vorzeichens aufteilt. Wählt man den Betrag des Feldindex groß gegen eins, so werden die Bewegungsgleichungen für beide Schwingungsrichtungen nahezu identisch; mit der Stabilität etwa der horizontalen Schwingungen ist dann auch die der vertikalen Schwingungen gewährleistet. Quantitativ ergibt sich in diesem Fall nach CLS für μ die folgende Gleichung

$$\cos\mu = \cos\pi\frac{n^{\frac{1}{2}}}{M}\cosh\pi\frac{n^{\frac{1}{2}}}{M}. \tag{E.15}$$

Dabei stellt n den absoluten Betrag des Feldindex dar. Man hat Stabilität, wenn die rechte Seite dem Betrag nach kleiner als eins bleibt. Als kennzeichnender Parameter tritt in der Stabilitätsbedingung die Größe $n^{\frac{1}{2}}/M$ auf.

Man kann leicht Varianten dieses Entwurfs angeben. Zum Beispiel können die Feldindizes aufeinander folgender Sektoren zwar weiterhin mit entgegengesetzten Vorzeichen aber von verschiedenem Betrag gewählt werden. Die entsprechenden Modifikationen der Gl. (E.14) sind von CLS angegeben worden. Es ist dann nötig, auf die Stabilität beider Schwingungsrichtungen besonders zu achten. Das von CLS angegebene Stabilitätsdiagramm ist (für den untersten Stabilitätsbereich) in Abb. 3 wiedergegeben. Die Bedeutung der im Innern eingezeichneten Linien wird in Unterabschnitt *b* in Zusammenhang mit der Diskussion von Feldfehlern erläutert werden. In entsprechender Weise wurden von verschiedenen Verfassern die Stabilitätsbedingungen bei verschiedener Länge der Sektoren und bei Einführung feldfreier Strecken ($\varrho=0$, $n=0$) zwischen den feldführenden Sektoren berechnet.

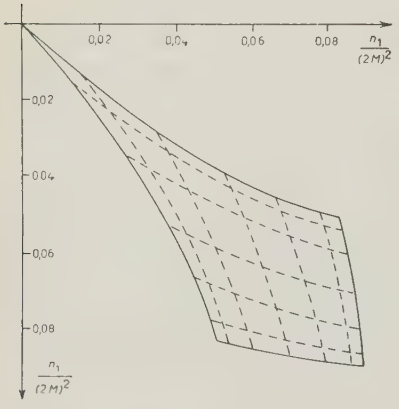


Abb. 3. – Stabilitätsdiagramm nach COURANT, LIVINGSTON und SNYDER für eine Anordnung, bei der jeder Periodizitätsbereich aus zwei Sektoren gleicher Länge mit Feldindizes $+n_1$ und $-n_2$ besteht (Grenzfall großen Wertes von M). Auf den gestrichelten Linien hat man Resonanzen mit ganz- oder halbzahligem Wert von Q für die vertikale oder horizontale Schwingungsrichtung (eingezeichnet für $M=6$). Instabilitätskurven für Kopplungsresonanzen sind nicht eingezeichnet.

b) *Typen von Feldfehlern und ihre Auswirkungen.* – Bei dem Bau eines Teilchenbeschleunigers treten unvermeidliche Abweichungen von dem idealen Entwurf auf. In dieser Arbeit soll untersucht werden, wie sich diese Abweichungen auf die Stabilität

der Betatronschwingungen auswirken und unter welchen Umständen sie diese Stabilität gefährden können. Die beiden Typen von Fehlern, die betrachtet werden sollen, sind

1. Fehler des magnetischen Führungsfeldes am Ort der idealen Gleichgewichtsbahn ([11]),
2. Fehler im Feldindex.

In der linearen Näherung der Bewegungsgleichungen kann ein weiterer Fehler dadurch entstehen, daß Gl. (E.4) nicht streng erfüllt ist ([12]). Er soll in dieser Arbeit nicht behandelt werden.

Als Ursachen für Fehler im Führungsfeld kommen u.a. in Frage: remanente Felder und Felder infolge von Wirbelströmen (insbesondere im Anfang des Beschleunigungszyklus) sowie Fehler in der Aufstellung der einzelnen Magnete in radialer und vertikaler Richtung. Derartige Aufstellungsfehler rufen wegen des inhomogenen Charakters des magnetischen Feldes eine Änderung des Führungsfeldes auf der ursprünglichen Gleichgewichtsbahn hervor. Ursachen für Fehler des Feldindex sind u.a. gegeben durch unterschiedliches Verhalten der einzelnen Magnete, insbesondere bei Sättigung, unterschiedliche Länge und Fehler in der Aufstellung der Magnete entlang der Gleichgewichtsbahn.

Charakteristisch für das Synchrotron ist, daß die zu beschleunigenden Teilchen viele Male in der Nähe einer geschlossenen Bahn herumgeführt werden und daß derselbe Fehler immer von neuem auf die Teilchen einwirkt. Die Stärke der Auswirkung der Fehler wird dabei ganz davon abhängen, wie sich die Phase der Betatronschwingungen während eines Umlaufs, also zwischen zwei Einwirkungen desselben Fehlers, ändert. Derartige Fehler sind bei dem neuen « Prinzip starker Stabilisierung » sehr viel bedeutungsvoller als bei dem älteren Stabilisierungsprinzip, denn hier kann eine große Anzahl von Betatronwellen auf dem Umfang liegen und daher schon eine geringe Änderung der Betriebsparameter zu einer wesentlichen Änderung der Differenz der Schwingungsphase zwischen zwei Umläufen führen. Es wird sich zeigen, daß es kritische Betriebsbedingungen in der Tat gibt. Ihre Existenz führt zu Grenzen, innerhalb deren die Betriebsparameter konstant zu halten sind. Läßt sich diese Konstanz nicht erreichen, so muß durch andere Mittel versucht werden, die kritische Auswirkung der Fehler zu vermeiden.

Feldfehler führen mathematisch zu Zusatzgliedern in der Bewegungsglei-

([11]) J. D. Lawson (unveröffentlichter CERN-Bericht) machte als erster darauf aufmerksam, daß derartige Fehler, speziell von Aufstellungsfehlern der Magnete herrührend, den Betrieb des Synchrotrons unmöglich machen können, falls eine ganze Zahl von Betatronwellenlängen auf dem Umfang liegt.

([12]) Auf diesen Fehler, der zu einer resonanzartigen Kopplung zwischen vertikaler und horizontaler Schwingung führen kann, wurde zuerst von P. A. Sturrock (unveröffentlichter CERN-Bericht) hingewiesen.

chung (E.6). Zunächst sollen Fehler im Führungsfeld betrachtet werden. Besonders einfach ist der Fall, wenn dieser Fehler durch fehlerhafte Aufstellung der Magnete in radialer Richtung hervorgerufen wird (die Beeinflussung der vertikalen Schwingungen erfolgt entsprechend durch Aufstellungsfehler in vertikaler Richtung). Ist der das Führungsfeld erzeugende Magnet am Ort ϑ um ein Stück $\delta\xi(\vartheta)$ radial senkrecht zur Gleichgewichtsbahn aus seiner idealen Stellung herausgerückt, so liegt der Mittelpunkt der rücktreibenden (oder forttreibenden) Kräfte an dieser Stelle bei $r = \delta\xi(\vartheta)$ statt bei $r = 0$. Statt Gl. (E.6) hat man daher jetzt folgende Bewegungsgleichung

$$r''(\vartheta) - n(\vartheta)(r(\vartheta) - \delta\xi(\vartheta)) = 0 \tag{E.16}$$

oder

$$r''(\vartheta) - n(\vartheta)r(\vartheta) = -n(\vartheta)\,\delta\xi(\vartheta)\,, \tag{E.17}$$

also eine inhomogene Differentialgleichung, die in geschlossener Form gelöst werden kann, falls zwei linear unabhängige Lösungen der zugehörigen homogenen Gleichung bekannt sind (siehe Teil I, Abschn. 1*a*). Ist, etwas allgemeiner, der Fehler im magnetischen Führungsfeld gegeben durch [13])

$$\delta H_z(\vartheta) = H_{z\,\mathrm{real}}(\vartheta) - H_{z\,\mathrm{ideal}}(\vartheta)\,, \tag{E.18}$$

so folgt aus der in Anh. 1 gegebenen Ableitung der Bewegungsgleichungen, daß Gl. (E.6) zu ersetzen ist durch

$$r''(\vartheta) - n(\vartheta)r(\vartheta) = \delta f(\vartheta) \tag{E.19}$$

mit

$$\delta f(\vartheta) = -\frac{R^2}{\varrho(\vartheta)}\,\frac{\delta H_z(\vartheta)}{H_z(\vartheta)} \tag{E.20}$$

$(H_z(\vartheta) \equiv H_{z\,\mathrm{ideal}}(\vartheta))$.

Entsprechend sei ein Fehler im Feldindex gegeben durch

$$\delta n(\vartheta) = n_{\mathrm{real}}(\vartheta) - n_{\mathrm{ideal}}(\vartheta)\,. \tag{E.21}$$

Die modifizierte Bewegungsgleichung (E.6) lautet dann

$$r''(\vartheta) - (n(\vartheta) + \delta n(\vartheta))r(\vartheta) = 0 \tag{E.22}$$

$(n(\vartheta) \equiv n_{\mathrm{ideal}}(\vartheta))$. Es entsteht also eine abgeänderte homogene Differentialgleichung. Das allgemeine Verhalten der Lösungen soll weiter unten in diesem

---

[13]) $H_{z\,\mathrm{ideal}}(\vartheta)$ ist durch Gl. (E.1) definiert, $H_{z\,\mathrm{real}}(\vartheta)$ stellt dagegen den Wert von $H_z(\vartheta)$ am Ort der idealen Gleichgewichtsbahn in dem realen, mit Fehlern behafteten Synchrotron dar.

Abschnitt diskutiert werden. Für die explizite Lösung soll später (Teil I, Abschn. 2) ein Näherungsverfahren verwendet werden. Im allgemeinen werden natürlich beide Fehler gleichzeitig — und dazu noch der hier fortgelassene Kopplungsfehler — auftreten. Aus methodischen Gründen ist es jedoch zweckmäßig, die Untersuchung der beiden Typen von Fehlern zunächst nach einander durchzuführen. In Teil I, Abschn. 3*a* wird die gleichzeitige Einwirkung beider Fehler diskutiert werden.

Es soll jetzt eine qualitative Diskussion der Auswirkung beider Typen von Fehlern gegeben werden. Bei den Fehlern im magnetischen Führungsfeld (Gl. (E.19)) läßt sich eine mechanische Analogie heranziehen. Hierzu sieht man von dem eigentlichen Charakteristikum des Prinzips starker Stabilisierung, nämlich der Abhängigkeit des Koeffizienten n von ϑ ab und betrachtet statt dessen Schwingungen, bei denen der Koeffizient in der rücktreibenden Kraft konstant ist. Ein System, das derartige Schwingungen ausführt, ist etwa durch eine aufgehängte Spiralfeder mit einem angehängten Gewicht gegeben. Eine Gleichung vom Typ (E.19) beschreibt dann Schwingungen dieses Systems unter Wirkung einer äußeren Kraft; speziell in Gl. (E.17) kann $\delta\xi(\vartheta)$ unmittelbar als äußere Verrückung des Aufhängepunkts der Spiralfeder interpretiert werden. Dem periodischen Durchlaufen immer derselben Fehler entspricht im Modell eine periodische äußere Kraft bzw. äußere Verrückung. Nun ist wohlbekannt, daß die Federschwingung unter der Einwirkung einer äußeren (nicht notwendig sinusförmigen) Kraft oder Verrückung dargestellt werden kann als Überlagerung einer freien Schwingung und einer erzwungenen Schwingung mit derselben Periode wie die Anregung. Die Amplitude der erzwungenen Schwingung geht (formal) gegen unendlich, wenn sich die Periode der (i.a. nicht sinusförmigen!) Anregung der Periode der freien Schwingung oder einem Vielfachen hiervon nähert. Zurückübersetzt in die Sprache der Betatronschwingungen bedeutet dies, daß jede Teilchenbahn unter dem Einfluß von Fehlern im Führungsfeld aufgefaßt werden kann als Superposition einer freien Betatronschwingung und einer nach einem Umlauf in sich geschlossenen Bahn; diese geschlossene Bahn kann als geänderte Gleichgewichtsbahn verstanden werden und tritt an die Stelle der idealen Gleichgewichtsbahn im störungsfreien Synchrotron. Nähert sich die Zahl der Betatronwellenlängen auf dem Umfang (die Größe Q aus Gl. (E.12)) einer ganzen Zahl, so geht die Abweichung der geänderten von der idealen Gleichgewichtsbahn (formal) gegen unendlich. Man hat dann eine resonanzartige Auswanderung des ganzen Strahls.

Auch die Auswirkung von Fehlern im Feldindex kann man sich am Beispiel der schwingenden Feder klarmachen. Fehlern im Feldindex, d.h. in der rücktreibenden Kraft, entspricht eine periodische Änderung der Federkonstanten, z.B. jeweils eine kurzzeitige Erweichung der Feder. Allgemeiner und unmittelbarer ist eine mathematische Diskussion der Gl. (E.22) mit den in Unterabschn. *a* bereitgestellten Hilfsmitteln. Der abgeänderte Feldindex $n(\vartheta) + \delta n(\vartheta)$

ist i.a. nicht mehr periodisch mit der Periode ϑ_0, wohl aber weiterhin mit 2π (= ein Umlauf). Auch in Gegenwart von Fehlern im Feldindex bleibt die Differentialgleichung für die Betatronschwingungen vom Floquetschen Typ, nur ist die Grundperiode jetzt 2π. Die Vernichtung der Periode ϑ_0 durch die Fehler führt dazu, daß instabile Bereiche auftreten können, wenn die Betatronschwingungen über einen Umlaut periodisch oder antiperiodisch sind, d.h. wenn die Größe Q aus Gl. (E.12) ganz- oder halbzahlig ist [14]. Ist der Fehler $\delta n(\vartheta)$ klein, so werden diese instabilen Bereiche schmal sein, d.h. sie werden bereits bei einer verhältnismäßig geringen Änderung der Betriebsparameter völlig überstrichen; außerdem werden sie in der Nähe solcher Betriebsbedingungen liegen, für die in dem ungestörten Beschleuniger [15] gerade eine ganze oder halbganze Zahl von Betatronschwingungen auf dem Umfang liegt. Die Instabilität äußert sich in einer monotonen Verbreiterung des Teilchenstrahls.

Nach Unterabschnitt *a* ändert sich die Gestalt von $r_1(\vartheta)$ bzw. $r_2(\vartheta)$ (Gl. (E.10)), wenn ein Stabilitätsbereich des idealen Synchrotrons monoton durchlaufen wird, von einer in ϑ_0 periodischen in eine in ϑ_0 antiperiodische Funktion (oder umgekehrt). Mit M (vgl. Gl. (E.8)) als Zahl der Perioden des Feldindex auf dem Umfang tritt bei einem solchen Durchlaufen des Stabilitätsbereichs $(M-1)$ mal (von den Grenzen des Stabilitätsbereichs abgesehen) der Fall ein, daß die Lösungen über einen vollen Umlauf periodisch oder antiperiodisch sind. Dann können sich jedesmal Fehler im Feldindex durch neue instabile Bereiche bemerkbar machen; jeder zweite derartige kritische Punkt (gezählt von demjenigen Rand des Stabilitätsbereichs, an dem man eine periodische Lösung hat) führt auch zu einer resonanzartigen Auswirkung der Fehler des Führungsfeldes, also zu einem unendlichen Anwachsen der Amplitude der geänderten Gleichgewichtsbahn. In einem zweidimensionalen Diagramm nach Art der Abb. 3 erhält man entsprechende kritische Kurven, von denen die eine Schar kritische Betriebsbedingungen für die vertikalen Schwingungen, die andere für die horizontalen Schwingungen angibt (gestrichelte Linien der Abb.). Kritische Kurven für Kopplungsresonanzen sind nicht eingetragen.

Die vorliegende Arbeit gliedert sich in zwei Teile. In Teil I werden die mathematischen Methoden für die Behandlung der Fehler entwickelt und allgemeine geschlossene Formeln für ihre Auswirkung angegeben. In Teil II erfolgt eine Auswertung dieser Formeln unter statistischen Voraussetzungen über die

---

[14] Für diese instabilen Bereiche wurde der Name « Stopbands » (Sperrbereiche) vorgeschlagen.

[15] Die Aufspaltung des wirklichen Feldindex in einen idealen (d.h. mit ϑ_0 periodischen) Anteil und eine Störung ist natürlich nicht eindeutig. Die Ergebnisse werden hierdurch jedoch nicht beeinflußt.

Fehler und eine Untersuchung der auftretenden Wahrscheinlichkeitsverteilungen. Bei der Planung eines Beschleunigers interessieren im wesentlichen derartige statistische Aussagen ([16]).

I. – Allgemeine Behandlung.

1. – Fehler im magnetischen Führungsfeld.

a) Gleichung der geänderten Gleichgewichtsbahn ([17]). – Das mathematische Problem der unter dem Einfluß von Fehlern im Führungsfeld geänderten Gleichgewichtsbahn besteht darin, solche Lösungen der Gl. (E.19) zu suchen, die periodisch mit der Periode 2π (ein Umlauf) sind

$$\hat{r}(\vartheta + 2\pi) = \hat{r}(\vartheta) \,. \tag{I.1.1}$$

Das ist in geschlossenener Form möglich. Denn die allgemeine Lösung der Gl. (E.19) kann angegeben werden, falls zwei linear unabhängige Lösungen $r_1(\vartheta)$, $r_2(\vartheta)$ (und daher die vollständige Lösung) der zugehörigen homogenen Gleichung (Gl. (E.6)) bekannt sind. Die allgemeine Lösung der inhomogenen Gleichung ist nämlich dann gegeben durch

$$r(\vartheta) = a_1 r_1(\vartheta) + a_2 r_2(\vartheta) + \frac{1}{W}\int_0^{\vartheta} \mathrm{d}\vartheta'\, \delta f(\vartheta')\{r_1(\vartheta')\, r_2(\vartheta) - r_2(\vartheta')\, r_1(\vartheta)\} \,. \tag{I.1.2}$$

Für die Ableitung folgt hieraus

$$r'(\vartheta) = a_1 r_1'(\vartheta) + a_2 r_2'(\vartheta) + \frac{1}{W}\int_0^{\vartheta} \mathrm{d}\vartheta'\, \delta f(\vartheta')\{r_1(\vartheta')\, r_2'(\vartheta) - r_2(\vartheta')\, r_1'(\vartheta)\} \,. \tag{I.1.3}$$

([16]) Ähnliche Untersuchungen wie die hier vorzulegenden sind an verschiedenen Stellen unabhängig von einander durchgeführt worden. Veröffentlichungen hierüber liegen kaum vor; die unveröffentlichten Berichte waren dem Verf. nur teilweise zugänglich. Eine Würdigung aller parallelen Entwicklungen war nicht möglich und wurde auch nicht beabsichtigt.

([17]) Die Bedeutung der Existenz einer geänderten Gleichgewichtsbahn in Gegenwart von Fehlern im Führungsfeld wurde dem Verf. Ende des Jahres 1952 deutlich; damals fanden in Kopenhagen im Zusammenhang mit den kritischen Bemerkungen von LAWSON (vgl. Fußn. ([12])) eine Reihe von Diskussionen statt, insbesondere mit Herrn T. SIGURGEIRSSON. Berechnungen und statistische Abschätzungen der geänderten Gleichgewichtsbahn wurden für ein vereinfachtes mathematisches Modell (konstantes, aber großes n) unabhängig u.a. von E. D. COURANT und E. BODENSTEDT durchgeführt.

Die Größe W bedeutet in diesen Gleichungen die durch (E.9) definierte Wronski determinante. Die Größen a_1 und a_2 sind beliebige Konstante.

Es ist zweckmäßig, für $r_1(\vartheta)$ und $r_2(\vartheta)$ die durch Gl. (E.10) definierten Floquetschen Lösungen zu verwenden, die einfache Periodizitätseigenschaften besitzen. Innerhalb eines Stabilitätsbereichs, also für reelles μ (Gl. (E.11)) kann man sie konjugiert komplex zu einander wählen; die Wronskideterminante ist dann rein imaginär ([18]). Für eine wirkliche Teilchenbahn, also reelles $r(\vartheta)$, sind dann die Konstanten a_1 und a_2 in Gln. (I.1.2), (I.1.3) konjugiert komplex zu einander. Wegen Gln. (E.11), (E.8) und (E.12), gilt für einen vollen Umlauf

$$\text{(I.1.4)} \qquad \begin{cases} r_1(\vartheta + 2\pi) = \exp[2\pi i Q] r_1(\vartheta) , \\ r_2(\vartheta + 2\pi) = \exp[-2\pi i Q] r_2(\vartheta) . \end{cases}$$

Gleichwertig mit der Periodizitätsbedingung Gl. (I.1.1) sind die beiden folgenden Bedingungen

$$\text{(I.1.5)} \qquad \hat{r}(2\pi) = \hat{r}(0) , \qquad \hat{r}'(2\pi) = \hat{r}'(0) .$$

Sie schreiben sich wegen Gln. (I.1.2), (I.1.3) folgendermaßen

$$\text{(I.1.6)} \qquad \begin{cases} \hat{a}_1(1 - \exp[-2\pi i Q]) r_1(2\pi) + \hat{a}_2(1 - \exp[+2\pi i Q]) r_2(2\pi) + \\ \qquad + \frac{1}{W} \int_0^{2\pi} \mathrm{d}\vartheta \, \delta f(\vartheta) \{ r_1(\vartheta) r_2(2\pi) - r_2(\vartheta) r_1(2\pi) \} = 0 , \\ \hat{a}_1(1 - \exp[-2\pi i Q]) r_1'(2\pi) + \hat{a}_2(1 - \exp[+2\pi i Q]) r_2'(2\pi) + \\ \qquad + \frac{1}{W} \int_0^{2\pi} \mathrm{d}\vartheta \, \delta f(\vartheta) \{ r_1(\vartheta) r_2'(2\pi) - r_2(\vartheta) r_1'(2\pi) \} = 0 . \end{cases}$$

Dieses Gleichungssystem ist nach den Konstanten $\hat{a}_1$ und $\hat{a}_2$ aufzulösen; dann stellt (I.1.2) die Gleichung der geänderten Gleichgewichtsbahn dar. Es ist aber zweckmäßig, Gl. (I.1.6) nicht als inhomogenes Gleichungssystem für $\hat{a}_1$ und $\hat{a}_2$ sondern als homogenes Gleichungssystem für die bei $r_1(2\pi)$ bzw. $r_1'(2\pi)$ und $r_2(2\pi)$ bzw. $r_2'(2\pi)$ stehenden Faktoren aufzufassen. Die Determinante dieses Gleichungssystems ist die Wronskideterminante, also ist sie von null verschieden

([18]) Aussagen über das Vorzeichen der Wronskideterminante werden in Anh. 2 gewonnen werden.

Damit hat man sofort

$$(\text{I.1.7}) \qquad a_1 = \frac{\int_0^{2\pi} \delta f(\vartheta)\, r_2(\vartheta)\, \mathrm{d}\vartheta}{W(1 - \exp[-2\pi i Q])}, \qquad a_2 = -\frac{\int_0^{2\pi} \delta f(\vartheta)\, r_1(\vartheta)\, \mathrm{d}\vartheta}{W(1 - \exp[+2\pi i Q])}.$$

Man sieht somit, daß die geänderte Gleichgewichtsbahn in der linearen Näherung eindeutig bestimmt ist, solange die Zahl Q der Betatronwellenlängen auf dem Umfang von einer ganzen Zahl verschieden ist.

Durch Einsetzen von Gl. (I.1.7) in Gl. (I.1.2) kann jetzt ein geschlossener Ausdruck für die geänderte Gleichgewichtsbahn gewonnen werde. Es ergibt sich

$$(\text{I.1.8}) \qquad \hat{r}(\vartheta) = \frac{1}{W} \int_{\vartheta - 2\pi}^{\vartheta} \mathrm{d}\vartheta'\, \delta f(\vartheta') \left\{ \frac{r_1(\vartheta')\, r_2(\vartheta)}{1 - \exp[-2\pi i Q]} - \frac{r_2(\vartheta')\, r_1(\vartheta)}{\exp[-2\pi i Q] - 1} \right\}.$$

Diese Gleichung wird am einfachsten gewonnen, indem man in Gln. (I.1.2) und (I.1.7) den Nullpunkt so verschiebt, daß der gerade betrachtete Ort auf der Gleichgewichtsbahn bei $\vartheta = 2\pi$ liegt, was keine Einschränkung der Allgemeinheit bedeutet.

Jede andere Teilchenbahn (andere Wahl der Konstanten a_1, a_2) führt freie Betatronschwingungen um diese geänderte Gleichgewichtsbahn herum aus. Untersuchungen über den Einfluß von Störungen im magnetischen Führungsfeld, hervorgerufen durch Aufstellungsfehler, wurden ohne Anwendung des Konzepts der geänderten Gleichgewichtsbahn von verschiedenen Autoren durchgeführt [19]. Natürlich zeigt sich auch dann, daß jede Betatronschwingung in der Amplitude beschränkt bleibt, falls Q nicht ganzzahlig ist. Verf. zieht jedoch aus methodischen Gründen den Begriff der geänderten Gleichgewichtsbahn vor; dieser wird sich auch bei der Untersuchung der gleichzeitigen Auswirkung von Fehlern im Führungsfeld und im Feldindex als nützlich erweisen.

Alle expliziten Überlegungen dieser Arbeit vernachlässigen den Einfluß der Beschleunigungsspannung und der Änderung der Betriebsparameter (des «idealen» Feldindex sowie aller Fehler) während des Beschleunigungszyklus. In Wahrheit liegt aber kein stationäres Problem vor. So bleibt die Lage der geänderten Gleichgewichtsbahn sicher nicht während des ganzen Beschleunigungszyklus dieselbe. Man kann jedoch zeigen, daß der Teilchenstrahl einer hinreichend langsamen Änderung der momentanen Gleichgewichtsbahn folgt.

Im Resonanzfall selbst (Q ganzzahlig) existiert nach Gl. (I.1.7) im allgemeinen — wenn nicht gerade die auftretenden Integrale verschwinden — keine

[19] U.a. E. Regenstreif (unveröffentlichte CERN-Berichte), J. S. Bell: *A.E.R.E.* T/R 1114; M. Sand und B. Touchek: *Nuovo Cimento*, **10**, 604 (1953).

geschlossene Bahn im Synchrotron. Alle Bahnen werden nach einem Umlauf um dasselbe $\delta r(\vartheta)$ verschoben; für dieses gilt nach Gl. (I.1.2)

$$\delta r(\vartheta) = \frac{1}{W}\int_{\vartheta-2\pi}^{\vartheta} \mathrm{d}\vartheta'\, \delta f(\vartheta')\{r_1(\vartheta')\, r_2(\vartheta) - r_2(\vartheta')\, r_1(\vartheta)\}\,. \tag{I.1.9}$$

Physikalisch ist das monotone Auswandern aller Teilchenbahnen sofort verständlich: Die Fehler wirken stets in derselben Phase auf die Teilchen ein und superponieren sich daher fortwährend statt sich im Lauf der Umläufe wieder zu kompensieren. Die Bedingung für das Unterbleiben der Auswanderung, nämlich das Verschwinden der in Gl. (I.1.9) auftretenden Integrale, bedeutet anschaulich, daß die Störungen an keiner Bahn über einen Umlauf Arbeit leisten.

b) Spezielle Fälle. – Ist der Fehler im Führungsfeld streng periodisch mit der ursprünglichen Periode ϑ_0

$$\delta f(\vartheta + \vartheta_0) = \delta f(\vartheta)\,, \tag{I.1.10}$$

so treten bei ganzzahligen Werten von Q keine Resonanzerscheinungen auf; das Teilchen «merkt» gar nicht, wann ein Umlauf vorüber ist. Dieser Fall ist allerdings nur von mathematischem Interesse, da sich eine derartige strenge Periodizität praktisch nicht verwirklichen läßt. Die geänderte Gleichgewichtsbahn ist nach Unterabschn. *a* eindeutig bestimmt und daher mit der in ϑ_0 periodischen Bahn identisch. Man hat lediglich ein etwas abgeändertes ideales Synchrotron. Mathematisch läßt sich die Behauptung der Resonanzfreiheit leicht aus der allgemeinen Theorie gewinnen. Für den Koeffizienten von $r_2(\vartheta)$ in Gl. (I.1.8) gilt nämlich wegen (I.1.10) und (E.10)

$$\frac{\int_{\vartheta-\vartheta_0}^{\vartheta} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta'}{W(1-\exp[-2\pi i Q])} \sum_{m=1}^{M} \exp[-i(m-1)\mu] = \frac{\int_{\vartheta-\vartheta_0}^{\vartheta} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta'}{W(1 \quad \exp[\quad i\mu])}\,. \tag{I.1.11}$$

Der zweite Spezialfall, der in diesem Unterabschnitt behandelt werden soll, bezieht sich auf Feldfehler, die allein durch Aufstellungsfehler nach Gl. (E.17) hervorgerufen werden; der Aufstellungsfehler $\delta\xi(\vartheta)$ soll dabei eine verglichen mit der Betatronwellenlänge langsam veränderliche Funktion darstellen. Im mechanischen Modell der schwingenden Feder bedeutet dies, daß die Lage des Aufhängepunktes langsam verglichen mit der Schwingungsperiode verändert wird. Dort ist das Resultat klar: das aufgehängte Gewicht wird der Bewegung

des Aufhängepunktes einfach folgen ([20]). Ebenso wird man erwarten, daß auch im Synchrotron mit starker Stabilisierung die Gleichgewichtsbahn solchen « adiabatischen » Aufstellungsfehlern einfach folgt. Wegen der komplizierteren Bewegungsgleichung und des komplizierteren Einflusses von Aufstellungsfehlern ist das jedoch nicht ganz selbstverständlich und soll besonders bewiesen werden.

Durch Vergleich der Gln. (E.17) und (E.19) folgt unter Heranziehung der Gl. (E.6) für die freien Betatronschwingungen zunächst

$$\int_{\vartheta-2\pi}^{\vartheta} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta' = -\int_{\vartheta-2\pi}^{\vartheta} \delta\xi(\vartheta')\, n(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta' = -\int_{\vartheta-2\pi}^{\vartheta} \delta\xi(\vartheta')\, r_1''(\vartheta')\, \mathrm{d}\vartheta' \,. \tag{I.1.12}$$

Nach zweimaliger partieller Integration ergibt sich hieraus

$$\int_{\vartheta-2\pi}^{\vartheta} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta' = \left(-\delta\xi(\vartheta) r_1'(\vartheta) + \delta\xi'(\vartheta) r_1(\vartheta)\right)\left(1-\exp\left[-2\pi i Q\right]\right) - \\ - \int_{\vartheta-2\pi}^{\vartheta} \delta\xi''(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta' \,. \tag{I.1.13}$$

Wenn sich $\delta\xi(\vartheta)$ langsam ändert verglichen mit $r_1(\vartheta)$, so wird das Integral auf der rechten Seite klein sein gegen das auf der linken, d.h. gegen die restlichen Glieder auf der rechten Seite. Einsetzen in die Gl. (I.1.8) für die geänderte Gleichgewichtsbahn und Beachtung der Definition (E.9) der Wronskideterminante liefert somit

$$r(\vartheta) = \delta\xi(\vartheta) - \frac{1}{W}\int_{\vartheta-2\pi}^{\vartheta} \mathrm{d}\vartheta'\, \delta\xi''(\vartheta') \left\{ \frac{r_1(\vartheta') r_2(\vartheta)}{1-\exp\left[-2\pi i Q\right]} - \frac{r_2(\vartheta') r_1(\vartheta)}{1-\exp\left[+2\pi i Q\right]} \right\}, \tag{I.1.14}$$

wobei die außer $\delta\xi(\vartheta)$ auf der rechten Seite stehenden Glieder einen Korrekturterm darstellen. Vielfach wird bei derartigen adiabatischen Aufstellungsfehlern auch das Vakuumrohr der Kurve $\delta\xi(\vartheta)$ folgen und die einzigen Glieder von praktischem Interesse sind diese Korrekturglieder.

c) Anwendung auf die « synchrone Bahn ». – In den vorangegangenen Unterabschnitten wurden nur solche Teilchen betrachtet, deren Energie bzw.

([20]) Dieses Ergebnis, formuliert für Betatronschwingungen, wurde von E. D. Courant (unveröffentlicht) aus seinen vereinfachten Rechnungen herausgelesen und ausgesprochen.

Impuls mit dem Magnetfeld des idealen Synchrotrons während des ganzen Beschleunigungszyklus durch Gl. (E.1) in Beziehung steht. Solche Teilchen würden in dem idealen Synchrotron unter idealen Betriebsbedingungen ständig auf der Gleichgewichtsbahn umlaufen. Tatsächlich aber treten im Synchrotron auch Teilchen mit einem Impulsfehler δp auf. Dieser Impulsfehler und der Fehler in der Phasenlage in Bezug auf die beschleunigende Wechselspannung führen Schwingungen um die Nullwerte aus ([3]). Teilchen mit (für den Augenblick als zeitlich konstant angenommenem) Impulsfehler besitzen eine von der idealen Gleichgewichtsbahn verschiedene Gleichgewichtsbahn. Diese Bahn wird als « synchrone Bahn » bezeichnet. Die Länge der synchronen Bahn ist von derjenigen der idealen Gleichgewichtsbahn verschieden. Mit dieser Wegverlängerung oder -verkürzung ist eine Änderung der Umlaufszeit verbunden; hierauf beruht der Mechanismus der oben erwähnten Impuls- und Phasenschwingungen. Natürlich liegt bei der Berechnung der synchronen Bahn eigentlich kein stationäres Problem vor. Infolge der Impulsschwingungen verändert sich nämlich die synchrone Bahn langsam (verglichen mit der Wellenlänge der Betatronschwingungen); diese oszillatorische Veränderung der synchronen Bahn wird als Synchrotronschwingung bezeichnet. Im Vakuumrohr muß in horizontaler Richtung hinreichend Platz für die Synchrotronschwingungen vorhanden sein.

Die synchrone Bahn soll jetzt, unter Vernachlässigung der langsamen Veränderung von δp, berechnet werden. Teilchen mit einem derartigen Impulsfehler finden auf der ursprünglichen Gleichgewichtsbahn nicht das ihrem Impuls entsprechende Magnetfeld vor (von eigentlichen Feldfehlern sei hierbei abgesehen); es weicht nach Gl. (E.1) von demjenigen, das sie auf der Gleichgewichtsbahn halten würde, ab um

$$\delta H_z(\vartheta) = -H_z(\vartheta)\frac{\delta p}{p}. \tag{I.1.15}$$

Die zugehörige Größe $\delta f(\vartheta)$ (Gl. (E.20)) ist für solche Teilchen daher gegeben durch

$$\delta f(\vartheta) = \frac{R^2}{\varrho(\vartheta)}\frac{\delta p}{p}. \tag{I.1.16}$$

Ist der Krümmungsradius $\varrho(\vartheta)$ streng periodisch mit ϑ_0, so lautet die Gleichung der synchronen Bahn für einen Impulsfehler δp folgendermaßen

$$r_{\text{synchr}}(\vartheta) = \frac{1}{W}\int_{\vartheta}^{\vartheta+\vartheta_0} d\vartheta' \frac{R}{\varrho(\vartheta')}\left\{\frac{r_1(\vartheta')r_2(\vartheta)}{1-\exp[-i\mu]} - \frac{r_2(\vartheta')r_1(\vartheta)}{1-\exp[i\mu]}\right\}\frac{\delta p}{p}. \tag{I.1.17}$$

Die relative Wegverlängerung ergibt sich nach einfachen geometrischen Überlegungen zu

$$\frac{\delta L}{L} = \frac{1}{\vartheta_0}\int_0^{\vartheta_0} \frac{r_{\text{synchr}}(\vartheta)}{\varrho(\vartheta)}\,\mathrm{d}\vartheta\,. \tag{I.1.18}$$

Nach Einsetzen von (I.1.17) folgt hieraus

$$\frac{\delta L}{L} = \frac{1}{W\vartheta_0}\int_0^{\vartheta_0}\mathrm{d}\vartheta \int_{\vartheta-\vartheta_0}^{\vartheta}\mathrm{d}\vartheta' \frac{R}{\varrho(\vartheta)}\frac{R}{\varrho(\vartheta')}\left\{\frac{r_1(\vartheta')r_2(\vartheta)}{1-\exp[-i\mu]} - \frac{r_2(\vartheta')r_1(\vartheta)}{1-\exp[+i\mu]}\right\}\cdot\frac{\delta p}{p} \tag{I.1.19}$$

und nach partieller Integration

$$\frac{\delta L}{L} = -\frac{2}{W\vartheta_0(1-\exp[i\mu])}\int_0^{\vartheta_0}\mathrm{d}\vartheta\,\frac{R}{\varrho(\vartheta)}r_1(\vartheta)\int_{\vartheta-\vartheta_0}^{\vartheta}\mathrm{d}\vartheta'\,\frac{R}{\varrho(\vartheta')}r_2(\vartheta')\cdot\frac{\delta p}{p} \equiv \varkappa\frac{\delta p}{p}\,. \tag{I.1.20}$$

Allein die Größe $\varkappa$ geht in die Theorie der Impuls- und Phasenschwingungen ein. Insbesondere ist die « Übergangsenergie » $E_{\text{üb}}$, bei der die Phasenschwingungen instabil werden, nach CLS für $\varkappa > 1$ gegeben durch

$$\frac{E_{\text{üb}}}{E_0} = \varkappa^{\frac{1}{2}}\,. \tag{I.1.21}$$

Hierbei ist $E_{\text{üb}}$ als Gesamtenergie (einschl. Ruhenergie) gemeint und E_0 stellt die Ruhenergie der Teilchen dar.

Es ist nach der gegebenen Ableitung klar, daß sich die Betatronschwingungen der synchronen Bahn einfach überlagern. Weicht außerdem das Führungsfeld auch für Teilchen mit $\delta p = 0$ von dem idealen Wert ab, so überlagert sich die synchrone Bahn der in Unterabschn. *a* berechneten geänderten Gleichgewichtsbahn.

2. – Fehler im Feldindex.

Das Auftreten instabiler Bereiche in der Nähe von ganz- oder halbzahligen Werten von Q, der Zahl der Betatronwellenlängen auf dem Umfang, wird durch die Überlegungen am Schluß von Unterabschn. *b* der Einleitung allgemein bewiesen. Durch Untersuchung spezieller Beispiele wurden diese Instabilitäten unabhängig vom Verf. durch Lundquist und Regenstreif [21] gefunden.

[21] S. Lundquist: *Phys. Rev.*, **91**, 981 (1953); E. Regenstreif, unveröffentlichte CERN-Berichte.

In diesem Abschnitt soll eine störungsmäßige Behandlung des Problems durchgeführt werden, die sich naturgemäß besonders für kleine Fehler des Feldindex und kleine, die strenge Periodizität störende, Aufstellungsfehler entlang der Gleichgewichtsbahn sowie Längenfehler der einzelnen Magnete eignet.

a) Störungsrechnung für Fehler im Feldindex. – Die Wirkung eines Fehlers im Feldindex wird, im Gegensatz zu derjenigen eines Fehlers im Führungsfeld, durch eine homogene statt einer inhomogenen Differentialgleichung (Gl. (E.22)) beschrieben. Eine entsprechend einfache geschlossene Lösung kann daher nicht angegeben werden. Natürlich ist das strenge Durchrechnen spezieller Probleme möglich; ein allgemeinerer Einblick wird jedoch durch eine Störungsrechnung gewonnen.

Es stellt sich die Aufgabe, für die Hillsche Differentialgleichung (E.22) Lösungen (« Floquetsche Lösungen ») $\tilde{r}_1(\vartheta)$ und $\tilde{r}_2(\vartheta)$ mit den Periodizitätseigenschaften

$$\text{(I.2.1)} \qquad \begin{cases} \tilde{r}_1(\vartheta + 2\pi) = \exp\left[2\pi i\tilde{Q}\right]\tilde{r}_1(\vartheta)\,, \\ \tilde{r}_2(\vartheta + 2\pi) = \exp\left[-2\pi i\tilde{Q}\right]\tilde{r}_2(\vartheta)\,, \end{cases}$$

zu finden; vgl. Gl. (E.10) und beachte, daß der Feldindex jetzt nur noch periodisch mit der Periode 2π ist. Es werde gesetzt (vgl. Gln. (E.11) und (E.12))

$$\text{(I.2.2)} \qquad \begin{cases} r_1(\vartheta) = v_1(\vartheta)\exp\left[i\vartheta Q\right], \qquad \tilde{r}_1(\vartheta) = \tilde{v}_1(\vartheta)\exp\left[i\vartheta\tilde{Q}\right], \\ r_2(\vartheta) = v_2(\vartheta)\exp\left[-i\vartheta Q\right], \qquad \tilde{r}_2(\vartheta) = \tilde{v}_2(\vartheta)\exp\left[-i\vartheta\tilde{Q}\right], \end{cases}$$

wobei sich also der linke Satz von Lösungen auf das ungestörte Problem (mit $\delta n(\vartheta) = 0$), der rechte dagegen auf das gestörte bezieht. Für die Funktion $v_1(\vartheta)$ gilt die Differentialgleichung

$$\text{(I.2.3)} \qquad v_1''(\vartheta) + 2iQv_1'(\vartheta) - \left(n(\vartheta) + Q^2\right)v_1(\vartheta) = 0$$

mit der Periodizitätsbedingung

$$\text{(I.2.4)} \qquad v_1(\vartheta - 2\pi) = v_1(\vartheta)\,,$$

für die Lösung $\tilde{v}_1(\vartheta)$ des gestörten Problems gilt eine entsprechende Gleichung

$$\text{(I.2.5)} \qquad \tilde{v}_1''(\vartheta) + 2i\tilde{Q}\tilde{v}_1'(\vartheta) - \left(n(\vartheta) + \delta n(\vartheta) + \tilde{Q}^2\right)\tilde{v}_1(\vartheta) = 0$$

und Periodizitätsbedingung

$$\text{(I.2.6)} \qquad \tilde{v}_1(\vartheta + 2\pi) = \tilde{v}_1(\vartheta)\,.$$

Entsprechende Gleichungen, mit jeweils umgekehrtem Vorzeichen von Q bzw. $\tilde{Q}$ gelten für $v_2(\vartheta)$ bzw. $\tilde{v}_2(\vartheta)$.

Es liegt nahe, in Analogie zur Schrödingerschen Störungsrechnung in der Quantentheorie zu setzen

$$\text{(I.2.7)} \qquad \begin{cases} \tilde{Q} = \tilde{Q}^{(0)} + \tilde{Q}^{(1)} + \dots , \\ \tilde{v}_1(\vartheta) = \tilde{v}^{(0)}(\vartheta) + \tilde{v}^{(1)}(\vartheta) - \dots , \end{cases}$$

und dann zu fordern

$$\text{(I.2.8)} \qquad \begin{cases} \tilde{v}_1^{(0)''}(\vartheta) + 2i\tilde{Q}^{(0)}\tilde{v}_1^{(0)'}(\vartheta) - (n(\vartheta) + \tilde{Q}^{(0)^2})\tilde{v}_1^{(0)}(\vartheta) = 0 , \\ \tilde{v}_1^{(1)''}(\vartheta) + 2i\tilde{Q}^{(0)}\tilde{v}_1^{(1)'}(\vartheta) - (n(\vartheta) + \tilde{Q}^{(0)^2})\tilde{v}_1^{(1)}(\vartheta) = \\ \qquad = -2i\tilde{Q}^{(1)}\tilde{v}_1^{(0)'}(\vartheta) + 2\tilde{Q}^{(0)}\tilde{Q}^{(1)}\tilde{v}_1^{(0)}(\vartheta) + \delta n(\vartheta)\tilde{v}_1^{(0)}(\vartheta) , \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \end{cases}$$

sowie

$$\text{(I.2.9)} \qquad \begin{cases} \tilde{v}_1^{(0)}(\vartheta - 2\pi) = \tilde{v}_1^{(0)}(\vartheta) , \\ \tilde{v}_1^{(1)}(\vartheta + 2\pi) = \tilde{v}_1^{(1)}(\vartheta) , \\ \dots \dots \dots \end{cases}$$

Falls Q weder ganz- noch halbzahlig ist, folgt aus den ersten Gln. (I.2.8) und (I.2.9), daß man setzen kann

$$\text{(I.2.10)} \qquad \tilde{Q}^{(0)} = Q , \qquad \tilde{v}_1^{(0)}(\vartheta) = v_1(\vartheta) .$$

Allerdings ist die Größe $\tilde{Q}$ nur bis auf eine beliebige additive ganze Zahl und auch das zugehörige $\tilde{v}_1(\vartheta)$ nur bis auf einen beliebigen konstanten Faktor bestimmt, aber die so berechneten Funktionen $\tilde{r}_1(\vartheta)$ unterscheiden sich dann nur um einen Zahlenfaktor. Bei ganz- oder halbzahligem Q werden die ersten Gln. (I.2.8) und (I.2.9) auch von $r_2(\vartheta)\exp[-2i\vartheta Q]$ erfüllt; es liegt dann also ein entartetes Problem vor.

Bei einer ganzen Zahl von Betatronwellenlängen auf dem Umfang soll zur Vereinfachung der mathematischen Behandlung die in Gl. (I.2.2) auftretende Größe Q so umdefiniert werden, daß $Q = 0$ gilt. Man hat dann einfach

$$\text{(I.2.11)} \qquad r_1(\vartheta) = v_1(\vartheta) , \qquad r_2(\vartheta) = v_2(\vartheta) ,$$

und die Entartung des Problems (im Sinne der Störungsrechnung) wird evident, da $v_1(\vartheta)$ und $v_2(\vartheta)$ derselben Differentialgleichung und derselben Periodizitätsbedingung gehorchen. Für halbzahliges Q soll ebenfalls formal $Q = 0$ gesetzt

und dafür die Periodizitätsbedingungen (I.2.4), (I.2.6), (I.2.9) durch die entsprechenden Antiperiodizitätsforderungen ersetzt werden. Aus der ersten Gl. (I.2.8) kann dann $\tilde{Q}^{(0)} = 0$ gefolgert werden; $\tilde{r}_1^{(0)}(\vartheta)$, die nullte Näherung des gestörten Problems, ist jetzt aber als Linearkombination von $r_1(\vartheta)$ und $r_2(\vartheta)$ anzusetzen

$$\tilde{r}_1^{(0)}(\vartheta) = a_{11}^{(0)} r_1(\vartheta) + a_{12}^{(0)} r_2(\vartheta) \,. \tag{I.2.12}$$

Das Verhältnis der Koeffizienten $a_{11}^{(0)}$ und $a_{12}^{(0)}$ kann, genau wie bei dem entsprechenden entarteten Problem der Schrödingerschen Störungsrechnung, erst aus der Gleichung für die erste Näherung bestimmt werden.

In der zweiten Gl. (I.2.8) (mit $Q^{(0)}=0$) ist die rechte Seite bekannt; die Lösung dieser Gleichung kann nach Gl. (I.1.2) in geschlossener Form angegeben werden

$$\tilde{r}_1^{(1)}(\vartheta) = a_{11}^{(1)} r_1(\vartheta) + a_{12}^{(1)} r_2(\vartheta) + \frac{1}{W} \int_0^{\vartheta} \mathrm{d}\vartheta' \{ 2i\tilde{Q}^{(1)} \tilde{r}_1^{(0)\prime}(\vartheta') + \delta n(\vartheta') \tilde{r}_1^{(0)}(\vartheta') \} \cdot \tag{I.2.13}$$

$$\cdot \{ r_1(\vartheta') r_2(\vartheta) - r_2(\vartheta') r_1(\vartheta) \} \,,$$

wobei sich die Koeffizienten $a_{11}^{(1)}$, $a_{12}^{(1)}$ dann erst aus der zweiten Näherung ergeben. Eine entsprechende Gleichung (mit Ersetzung von $r_1(\vartheta)$, $r_2(\vartheta)$ durch ihre ersten Ableitungen) gilt nach Gl. (I.1.3) für die erste Ableitung $\tilde{r}_1^{(1)\prime}(\vartheta)$. Jetzt ist die Periodizitäts- (oder Antiperiodizitäts-) bedingung (I.2.9) heranzuziehen. Es müssen die in der Gleichung für $\tilde{r}_1^{(1)}(\vartheta)$ (Gl. (I.2.13)) und $\tilde{r}_1^{(1)\prime}(\vartheta)$ stehenden Integrale verschwinden, wenn sie von 0 bis 2π erstreckt werden. In derselben Weise, wie Gl. (I.1.7) aus Gl. (I.1.6) gewonnen wurde, folgt dann

$$\int_0^{2\pi} (-2i\tilde{Q}^{(1)} \tilde{r}_1^{(0)\prime}(\vartheta) + \delta n(\vartheta) \tilde{r}_1^{(0)}(\vartheta)) \begin{Bmatrix} r_1(\vartheta) \\ r_2(\vartheta) \end{Bmatrix} \mathrm{d}\vartheta = 0 \,. \tag{I.2.14}$$

Dies ist das Gleichungssystem, aus dem nach Einsetzen von Gl. (I.2.12) die Größe $\tilde{Q}^{(1)}$ und das Verhältnis der Koeffizienten $a_{11}^{(0)}$ und $a_{12}^{(0)}$ zu bestimmen sind.

Es werde jetzt die Abkürzung

$$N_{jk} = \int_0^{2\pi} \delta n(\vartheta) r_j(\vartheta) r_k(\vartheta) \,\mathrm{d}\vartheta \,, \tag{I.2.15}$$

$(j, k = 1, 2)$ eingeführt. Offenbar ist diese Größe symmetrisch in den Indizes j, k. Ferner gilt für konjugiert komplexe Funktionen $r_1(\vartheta)$, $r_2(\vartheta)$ (vgl. Abschn. 1*a*)

$$N_{11} = N_{22}^* \,, \qquad N_{12} = N_{12}^* = N_{21} \,. \tag{I.2.16}$$

Unter Benutzung der Definition (E.9) der Wronskideterminante und unter Beachtung der Periodizität (bzw. Antiperiodizität) von $r_1(\vartheta)$, $r_2(\vartheta)$ zeigt man außerdem leicht

$$\text{(I.2.17)} \qquad \begin{cases} \int_0^{2\pi} r_1(\vartheta)\, r_2'(\vartheta)\, \mathrm{d}\vartheta = -\int_0^{2\pi} r_2(\vartheta)\, r_1'(\vartheta)\, \mathrm{d}\vartheta = \pi W\,, \\ \int_0^{2\pi} r_1(\vartheta)\, r_1'(\vartheta)\, \mathrm{d}\vartheta = \int_0^{2\pi} r_2(\vartheta)\, r_2'(\vartheta)\, \mathrm{d}\vartheta = 0\,. \end{cases}$$

Somit schreibt sich das aus Gl. (I.2.14) folgende Gleichungssystem für die Koeffizienten $a_{11}^{(0)}$, $a_{12}^{(0)}$, das gleichzeitig die Änderung $\tilde{Q}^{(1)}$ der Zahl der Betatronwellenlängen auf dem Umfang in niedrigster Näherung zu berechnen erlaubt, in der Form

$$\text{(I.2.18)} \qquad a_{11}^{(0)}(N_{11} + 2\pi i W \tilde{Q}^{1}) + a_{12}^{(0)} N_{22} = 0\,, \qquad a_{11}^{(0)} N_{11} + a_{12}^{(0)}(N_{21} - 2\pi i W \tilde{Q}^{1}) = 0\,.$$

Die Determinante dieses Gleichungssystems muß verschwinden; unter Beachtung der Symmetrie von N_{jk} erhält man damit für $\tilde{Q}^{(1)}$ die Gleichung

$$\text{(I.2.19)} \qquad (Q^{(1)})^2 = -\frac{\det N}{(2\pi i W)^2}$$

mit

$$\text{(I.2.20)} \qquad \det N = N_{11} N_{22} - N_{12} N_{21}\,.$$

Man hat in dieser Näherung Stabilität für reelles $\tilde{Q}^{(1)}$ (oder $(\tilde{Q}^{(1)})^2 > 0$) und Instabilität für imaginäres $\tilde{Q}^{(1)}$ (oder $(\tilde{Q}^{(1)})^2 < 0$). Da die Wronskideterminante nach Abschn. 1*a* eine imaginäre Größe darstellt, hat man Stabilität für $\det N < 0$ und Instabilität für $\det N > 0$. Eine genauere Diskussion der Folgerungen aus dieser Gleichung erfolgt im nächsten Unterabschnitt.

Einige weitere Anwendungen der hier skizzierten Störungsrechnung werden in Anh. 2 gebracht.

b) *Gleichungen für den instabilen Bereich.* – Die Grundlage für die Untersuchung der durch Fehler im Feldindex hervorgerufenen Instabilität bildet Gl. (I.2.20). Wir zeigen zunächst, daß eine Instabilität nicht auftritt, wenn $\delta n(\vartheta)$ periodisch mit der ursprünglichen Periode ϑ_0 ist. Physikalisch leuchtet das sofort ein, da man durch eine solche Änderung des Feldindex (wenigstens, wenn sie hinreichend klein ist) nur einen abgeänderten idealen Beschleuniger erhält, in dem sich die Periode 2π in keiner Weise bemerkbar macht. Mathematisch ergibt sich dieses Resultat durch Berechnung von N_{11}. Es gilt näm-

lich (vgl. die entsprechenden Überlegungen am Anfang von Abschn. 1*b*)

$$\text{(I.2.21)} \qquad \begin{cases} N_{11} = \int\limits_0^{\vartheta_0} \delta n(\vartheta)(r_1(\vartheta))^2 \,\mathrm{d}\vartheta \cdot \sum\limits_{m=1}^{M} \exp[2i(m-1)\mu] = \\ \qquad = \int\limits_0^{\vartheta_0} \delta n(\vartheta)(r_1(\vartheta))^2 \,\mathrm{d}\vartheta \cdot \dfrac{\exp[4\pi i Q] - 1}{\exp[2i\mu] - 1} \,, \end{cases}$$

und das verschwindet für ganz- und halbzahligen Wert von Q. Nach Gl. (I.2.20) hat man dann stets

$$\text{(I.2.22)} \qquad \det N = -(N_{12})^2 \leqq 0 \,.$$

also Stabilität.

Zwecks Untersuchung der durch Fehler im Feldindex hervorgerufenen instabilen Bereiche soll der ungestörte Feldindex $n(\vartheta)$ in zwei Anteile zerlegt werden, einen Anteil, der exakt zu ganz- oder halbzahligem Q führt und ein Zusatzglied $\delta n_{\text{ideal}}(\vartheta)$, das die Änderung des idealen Arbeitspunktes bei Überstreichen einer kleinen Umgebung dieser Resonanzbedingungen wiedergibt. Formal damit gleichwertig ist, die Größe $n(\vartheta)$ in Gl. (E.22) als Feldindex, der zu ganz- oder halbzahligem Wert von Q führt, zu verstehen, und das Zusatzglied aufzuspalten in zwei Anteile

$$\text{(I.2.23)} \qquad \delta n(\vartheta) = \delta n_{\text{ideal}}(\vartheta) + \delta \hat{n}(\vartheta) \,,$$

wobei $\delta n_{\text{ideal}}(\vartheta)$ die Änderung des idealen, streng mit ϑ_0 periodischen Feldindex, $\delta \hat{n}(\vartheta)$ aber den eigentlichen Fehler bedeutet. Genau genommen wird der Fehler $\delta \hat{n}(\vartheta)$ auch von der Wahl des Arbeitspunktes, d.h. von $\delta n_{\text{ideal}}(\vartheta)$ abhängen; jedoch dürfte es sich hierbei um einen Effekt höherer Ordnung handeln.

Auf diese Weise spaltet N_{jk} (Gl. (I.2.15)) in eine Summe von zwei Termen auf, wobei zunächst der von $\delta n_{\text{ideal}}(\vartheta)$ herrührende Anteil untersucht werden soll. Es soll gezeigt werden, daß gilt

$$\text{(I.2.24)} \qquad \begin{cases} \int\limits_0^{2\pi} \delta n_{\text{ideal}}(\vartheta)(r_1(\vartheta))^2 \,\mathrm{d}\vartheta = \int\limits_0^{2\pi} \delta n_{\text{ideal}}(\vartheta)(r_2(\vartheta))^2 \,\mathrm{d}\vartheta = 0 \,, \\ \int\limits_0^{2\pi} \delta n_{\text{ideal}}(\vartheta)\, r_1(\vartheta)\, r_2(\vartheta) \,\mathrm{d}\vartheta = -2\pi i W \,\delta Q^{(1)} \,, \end{cases}$$

wobei $\delta Q^{(1)}$ die Änderung des Wertes von Q für den fehlerfreien Beschleuniger (in erster Näherung) bei Änderung von $n(\vartheta)$ um $\delta n_{\text{ideal}}(\vartheta)$ bedeutet. Für

$\delta n(\vartheta)$ $\delta n_{\text{ideal}}(\vartheta)$ muß nämlich $\tilde{r}_1(\vartheta)$ im Grenzfall $\delta n(\vartheta) \to 0$ in $r_1(\vartheta)$ übergehen, d.h. Gl. (I.2.18) muß richtig sein für $a_{12}^{(0)} = 0$. Das führt aber in der Tat zu der Behauptung (I.2.24). Die erste dieser beiden Gleichungen hätte auch aus den Periodizitätsbetrachtungen am Anfang dieses Unterabschnitts gewonnen werden können; für die zweite Gleichung wird in Anh. 2*a* ein unabhängiger Beweis gegeben werden.

Führt man jetzt die Bezeichnung

$$\hat{N}_{jk} = \int_0^{2\pi} \delta\hat{n}(\vartheta)\, r_j(\vartheta)\, r_k(\vartheta)\, \mathrm{d}\vartheta \tag{I.2.25}$$

ein, so folgt für die in Unterabschn. *a* definierten Größen N_{jk} in leicht verständlicher Matrixschreibweise

$$\begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix} = -2\pi i W\, \delta Q^{(1)} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} \hat{N}_{11} & \hat{N}_{12} \\ \hat{N}_{21} & \hat{N}_{22} \end{pmatrix}, \tag{I.2.26}$$

und für $\det N$ erhält man

$$\det N = -(2\pi i W)^2 \left\{ \left(\delta Q^{(1)} - \frac{\hat{N}_{12}}{2\pi i W} \right)^2 - \frac{|\hat{N}_{11}|^2}{(2\pi i W)^2} \right\} \tag{I.2.27}$$

und somit nach Gl. (I.2.19)

$$Q^{(1)2} = \left(\delta Q^{(1)} - \frac{N_{12}}{2\pi i W} \right)^2 - \left| \frac{N_{11}}{2\pi i W} \right|^2 . \tag{I.2.28}$$

Aus dieser Gleichung lassen sich in Verbindung mit Gln. (I.2.18) und (I.2.12) alle erforderlichen Auskünfte über die in der Nähe ganz- oder halbzahliger Werte von Q liegenden instabilen Bereiche — wenigstens in niedrigster störungstheoretischer Näherung — gewinnen. Es soll noch einmal wiederholt werden, daß $\delta Q^{(1)}$ die Differenz zwischen dem Wert von Q in dem zugeordneten idealen Synchrotron und der nächsten ganzen oder halbganzen Zahl bedeutet, $\tilde{Q}^{(1)}$ aber dadurch definiert ist, daß gemäß Gl. (I.2.1) in niedrigster störungstheoretischer Näherung gilt

$$\begin{cases} \tilde{r}_1(\vartheta + 2\pi) = \pm \exp\,[2\pi i \tilde{Q}^{(1)}] \tilde{r}_1(\vartheta)\,, \\ \tilde{r}_2(\vartheta + 2\pi) = \pm \exp\,[-2\pi i \tilde{Q}^{(1)}] \tilde{r}_2(\vartheta)\,. \end{cases} \tag{I.2.29}$$

Das obere Vorzeichen gilt in der Umgebung ganzzahligen, das untere in der Umgebung halbzahligen Wertes von Q. Man hat Stabilität für reelles, insta-

bilität für imaginäres $\tilde{Q}^{(1)}$. Die Grenzen des instabilen Bereichs liegen bei $\tilde{Q}^{(1)} = 0$. In Abb. 4 ist $\tilde{Q}^{(1)^2}$ als Funktion von $\delta Q^{(1)}$ aufgezeichnet. Die Breite des instabilen Bereichs ist gegeben durch

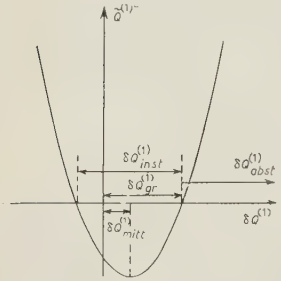


Abb. 4. – $\tilde{Q}^{(1)^2}$ nach Gl. (I.2.28) mit Einzeichnung einiger im Text auftretender Größen.

$$\text{(I.2.30)} \qquad \delta Q^{(1)}_{\text{inst}} = \frac{\dot{N}_{11}}{\pi W} = \left|\frac{\dot{N}_{22}}{\pi W}\right| ;$$

man hat — in dieser Näherung — keinen instabilen Bereich, wenn $\dot{N}_{11}$ verschwindet. Der Abstand des Mittelpunktes des instabilen Bereichs vom nächsten ganz- oder halbzahligen Wert von Q ergibt sich zu

$$\text{(I.2.31)} \qquad \delta Q^{(1)}_{\text{mitt}} = \frac{\dot{N}_{12}}{2\pi i W} .$$

Das ganze Störungsverfahren ist natürlich nur sinnvoll für Werte von $\delta Q^{(1)}$, die klein gegen eins sind.

Es ist innerhalb gewisser Grenzen willkürlich, was man bei einem gegebenen fehlerhaften Beschleuniger den zugeordneten fehlerfreien, idealen Beschleuniger nennt. Die Ergebnisse, insbesondere die in Gl. (I.2.30), (I.2.31) enthaltenen Aussagen, werden hierdurch jedoch nicht beeinflußt. Bei einer Umdefinition von $u_{\text{ideal}}(\vartheta)$, d.h. bei einer geänderten Aufspaltung (I.2.23), wird Gl. (I.2.30) überhaupt nicht verändert; $\delta Q^{(1)}_{\text{mitt}}$ ändert sich aber gerade in solcher Weise, daß die objektive Lage des instabilen Bereichs dieselbe bleibt.

c) Flanken der instabilen Bereiche. – Arbeitet man unter instabilen Betriebsbedingungen, also innerhalb eines der instabilen Bereiche, so gibt es eine spezielle Bahn, deren Amplitude während der Umläufe monoton abnimmt, und eine andere, deren Amplitude monoton zunimmt. Die allgemeine Teilchenbahn ist eine Linearkombination dieser beiden speziellen Lösungen der Bewegungsgleichung, wobei sich im Verlauf der Umläufe die Lösung mit zunehmender Amplitude durchsetzt. Arbeitet man zwar im stabilen Gebiet jedoch dicht an der Grenze eines instabilen Bereichs, so wird man ein schwebungsartiges Schwanken der Amplitude jeder Betatronschwingung erwarten, denn eine Zeit lang werden die Fehler des Feldindex etwa in solcher Weise wirken, daß die Amplitude zunimmt; dann jedoch verschiebt sich die Phasenlage der Schwingung relativ zu den Fehlern und sie wirken in umgekehrter, die Amplitude vermindernder, Richtung; anschließend wird wiederum eine Amplitudenzunahme erfolgen, und so fort. Die Größe dieser Schwebungen kann einem verbieten, zu nahe an den Grenzen eines instabilen Bereichs zu arbeiten. Man

kann von Flanken des instabilen Bereichs sprechen, die sich in den stabilen Bereich hinein erstrecken.

Diese qualitativen Überlegungen sollen jetzt quantitativ verschärft werden [22]. Hierzu ist die in den vorangehenden Unterabschnitten im Prinzip ermittelte Gestalt der Bahnen in der Nähe ganz- oder halbzahligen Wertes von Q genauer zu untersuchen.

Bei dem idealen, fehlerfreien Beschleuniger kann jede Betatronschwingung in der Form

$$r(\vartheta) = \xi r_1(\vartheta) + \xi^* r_2(\vartheta) \tag{I.2.32}$$

dargestellt werden ($r_1(\vartheta)$ und $r_2(\vartheta)$ sind Floquetsche Lösungen nach Gl. (E.10); es sei $r_2(\vartheta) = r_1^*(\vartheta)$). Hierbei seien ξ und ξ^* zueinander konjugiert komplexe Konstante. Schränkt man die Variable ϑ auf der rechten Seite der Gleichung (aber natürlich nicht auf der linken Seite!) etwa auf das Intervall $0 \leqslant \vartheta < 2\pi$ ein, so multiplizieren sich ξ bzw. ξ^* bei jedem Umlauf mit $\exp[2\pi i Q]$ bzw. $\exp[-2\pi i Q]$; diese Größen springen also bei unverändertem Betrag in der Phase. Es soll angenommen werden, daß die maximale Amplitude einer Betatronschwingung unabhängig von den speziellen Anfangsbedingungen proportional zu $|\xi|$ ist.

In einem Beschleuniger mit Fehlern im Feldindex gilt an Stelle von Gl. (I.2.32)

$$r(\vartheta) = \eta \tilde{r}_1(\vartheta) + \eta^* \tilde{r}_2(\vartheta) \tag{I.2.33}$$

mit $\tilde{r}_1(\vartheta)$ und $\tilde{r}_2(\vartheta)$ nach Gl. (I.2.1). Beschränkt man auch hier die Größe ϑ auf der rechten Seite auf das Intervall $0 \leqslant \vartheta < 2\pi$, so springt die Phase der Größen η bzw. η^* mit jedem Umlauf um $2\pi\tilde{Q}$ bzw. $-2\pi\tilde{Q}$. Nach Gl. (I.2.12) gilt nun näherungsweise (für $0 \leqslant \vartheta < 2\pi$)

$$\begin{cases} \tilde{r}_1(\vartheta) = a_1 r_1(\vartheta) + a_2 r_2(\vartheta)\,, \\ \tilde{r}_2(\vartheta) = a_2^* r_1(\vartheta) + a_1^* r_1(\vartheta)\,, \end{cases} \tag{I.2.34}$$

wobei die Bezeichnung der Koeffizienten durch Fortlassung von Indizes etwas vereinfacht wurde. In einem hinreichend kurzen Intervall der Größe ϑ ist daher immer noch eine Darstellung der Bahn in der ursprünglichen Form (I.2.32) möglich, wobei der momentane Wert der Größe ξ gegeben ist durch

$$\xi = \eta a_1 + \eta^* a_2^*\,. \tag{I.2.35}$$

---

[22] Ähnliche Rechnungen wurden unabhängig von J. S. Bell: *A.E.R.E. T/M* 86, durchgeführt. Eine vereinfachende Umformung des Resultats verdankt Verf. einer Korrespondenz mit Herrn Bell.

Da sich die Phase von η, η^* (nicht jedoch der Betrag) während der Umläufe ändert, schwankt $|\xi|$ schwebungsartig zwischen einem Maximalwert $|\xi|_{\max}$ und einem Minimalwert $|\xi|_{\min}$

$$|\xi|_{\max} = |\eta|(|a_1| + |a_2|), \quad |\xi|_{\min} = |\eta| \cdot ||a_1| - |a_2|| ; \tag{I.2.36}$$

und für das Verhältnis beider Werte, das als Schwebungsfaktor F bezeichnet sei, erhält man

$$F = \frac{|\xi|_{\max}}{|\xi|_{\min}} = \frac{1 + |a_2/a_1|}{|1 - |a_2/a_1||} . \tag{I.2.37}$$

Nach der oben ausgesprochenen Annahme gibt F zugleich das Verhältnis zwischen größter und kleinster Amplitude — jeweils verstanden als Maximalamplitude über eine größere Zahl von Betatronwellenlängen — wieder.

Das Verhältnis $|a_2/a_1|$ ist aus Gl. (I.2.18) zu entnehmen; man findet

$$\left|\frac{a_2}{a_1}\right| = \frac{|N_{12} + 2\pi i W \tilde{Q}^{(1)}|}{|N_{22}|} , \tag{I.2.38}$$

wobei $\tilde{Q}^{(1)}$ positiv gewählt sei. Die Größen N_{12} und N_{22} sollen in dieser Gleichung durch andere, den instabilen Bereich kennzeichnende Größen ersetzt werden. Zunächst gilt nach Gln. (I.2.26) und (I.2.31)

$$N_{12} = \overset{\triangle}{N}_{12} - 2\pi i W \delta Q^{(1)} = 2\pi i W (\delta Q^{(1)}_{\text{mitt}} - \delta Q^{(1)}) . \tag{I.2.39}$$

Ferner werde statt $\delta Q^{(1)}$ der Abstand $\delta Q^{(1)}_{\text{abst}}$ von der nächsten Grenze eines instabilen Bereichs eingeführt, für den offenbar gilt [23])

$$\delta Q^{(1)} = \delta Q^{(1)}_{\text{mitt}} + \tfrac{1}{2}\delta Q^{(1)}_{\text{inst}} + \delta Q^{(1)}_{\text{abst}} . \tag{I.2.40}$$

Verwendet man im Nenner von Gl. (I.2.38) noch Gl. (I.2.30), so erhält man

$$\left|\frac{a_2}{a_1}\right| = \frac{\delta Q^{(1)}_{\text{inst}} + 2(\delta Q^{(1)}_{\text{abst}} + \tilde{Q}^{(1)})}{\delta Q^{(1)}_{\text{inst}}} . \tag{I.2.41}$$

Für den Schwebungsfaktor folgt damit

$$F = \frac{\delta Q^{(1)}_{\text{inst}} + \delta Q^{(1)}_{\text{abst}} + \tilde{Q}^{(1)}}{\delta Q^{(1)}_{\text{abst}} + \tilde{Q}^{(1)}} . \tag{I.2.42}$$

[23]) Die ist zunächst richtig, falls man sich oberhalb und nicht unterhalb eines instabilen Bereichs befindet. Der Leser überzeugt sich jedoch leicht, daß Gl. (I.2.42) richtig bleibt, wenn man sich unterhalb eines instabilen Bereichs befindet; alle auftretenden Größen sind dann durch ihre Absolutbeträge zu ersetzen.

Schließlich möge noch die Größe $\tilde{Q}^{(1)}$ mittels der aus Gl. (I.2.28) folgenden Beziehung

$$\tilde{Q}^{(1)^2} = \delta Q^{(1)}_{\text{inst}}\, \delta Q^{(1)}_{\text{abst}} + (\delta Q^{(1)}_{\text{abst}})^2 \tag{I.2.43}$$

eliminiert werden. Nach einigen Umformungen erhält man dann

$$F = \sqrt{\frac{\delta Q^{(1)}_{\text{inst}}}{\delta Q^{(1)}_{\text{abst}}} + 1} \approx \sqrt{\frac{\delta Q^{(1)}_{\text{inst}}}{\delta Q^{(1)}_{\text{abst}}}}\,, \tag{I.2.44}$$

wobei der rechts stehende Näherungsausdruck in umittelbarer Nähe des instabilen Bereichs gültig ist. Der Schwebungsfaktor F ist in Abb. 5 dargestellt. Er geht an der Grenze des instabilen Bereichs gegen unendlich, wie zu erwarten war. Die Umgebung des instabilen Bereichs in der eine wesentliche Aufblähung des Strahls erfolgt, ergibt sich als klein gegen die Breite des instabilen Bereichs selbst. Die Flanken spielen daher praktisch keine große Rolle.

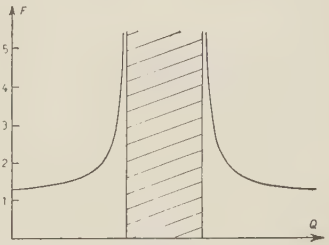


Abb. 5. – Der Schwebungsfaktor F in der Umgebung eines instabilen Bereichs nach Gl. (I.2.44).

Von größerem praktischen Interesse als der hier berechnete (« statische ») Schwebungsfaktor ist das Verhältnis der maximalen Amplitude, nicht zur minimalen Amplitude , sondern zur Amplitude der Betatronschwingungen im Augenblick der Injektion. Nimmt man an, daß die Injektion in hinreichendem Abstand von den benachbarten instabilen Bereichen erfolgt und daß man sich adiabatisch der Grenze eines solchen instabilen Bereichs nähert, so ergibt sich dieser « dynamische » Schwebungsfaktor als Quadratwurzel aus dem oben berechneten statischen. Dabei ist die « adiabatische Dämpfung » (vgl. Anh. 1*b*) unberücksichtigt; sie führt zu einer Verringerung aller Schwingungsamplituden im Verlauf des Beschleunigungsprozesses.

Die Zahl k der Umläufe zwischen zwei Schwebungsmaxima folgt aus der Bedingung, daß gelten muß

$$2k\tilde{Q}^{(1)} = 1\,. \tag{I.2.45}$$

In umittelbarer Nähe eines instabilen Bereichs hat man daher nach Gl. (I.2.43)

$$k = \frac{1}{2\sqrt{\delta Q^{(1)}_{\text{inst}}\, \delta Q^{(1)}_{\text{abst}}}}\,. \tag{I.2.46}$$

3. – Verschiedene Probleme.

a) *Gleichzeitige Wirkung von Fehlern im Führungsfeld und im Feldindex.* – Im allgemeinen werden beide in den Abschnitten 1 und 2 getrennt analysierten Typen von Fehlern zusammen auftreten; man hat dann sowohl eine geänderte Gleichgewichtsbahn als auch instabile Bereiche. Die Betatronschwingungen um die geänderte Gleichgewichtsbahn werden durch die Fehler im Führungsfeld nicht beeinflußt. Für sie gelten also weiterhin die Überlegungen des Abschn. 2. Auf die Gestalt der geänderten Gleichgewichtsbahn wirken sich dagegen beide Typen von Fehlern aus.

Zunächst möge beachtet werden, daß bei Ableitung der Gl. (I.1.8) für die geänderte Gleichgewichtsbahn zwar von der Periodizitätsbedingung (I.1.4) für einen vollen Umlauf, nicht aber von der stärkeren Periodizitätsbedingung (E.10) Gebrauch gemacht wurde. Gl. (I.1.8) gilt daher auch für ein Synchrotron, in dem außer Fehlern im Führungsfeld solche im Feldindex auftreten, nur sind dann statt der Floquetlösungen $r_1(\vartheta)$, $r_2(\vartheta)$ des fehlerfreien Beschleunigers die Lösungen $\tilde{r}_1(\vartheta)$, $\tilde{r}_2(\vartheta)$ gemäß Gl. (I.2.1) zu benutzen; entsprechend ist die Wronskideterminante W durch die aus diesen Lösungen gebildete Wronskideterminante

$$\tilde{W} = \tilde{r}_1(\vartheta)\tilde{r}_2'(\vartheta) - \tilde{r}_2(\vartheta)\tilde{r}_1'(\vartheta) \tag{I.3.1}$$

zu ersetzen. Man erhält damit folgende Gleichung für die geänderte Gleichgewichtsbahn unter dem Einfluß beider Typen von Fehlern

$$\hat{r}(\vartheta) = \frac{1}{\tilde{W}} \int_{\vartheta-2\pi}^{\vartheta} \mathrm{d}\vartheta' \, \delta f(\vartheta') \left\{ \frac{\tilde{r}_1(\vartheta')\tilde{r}_2(\vartheta)}{1 - \exp[-2\pi i \tilde{Q}]} - \frac{\tilde{r}_2(\vartheta')\tilde{r}_1(\vartheta)}{1 - \exp[+2\pi i \tilde{Q}]} \right\}. \tag{I.3.2}$$

Wie man sieht, tritt Resonanz, d.h. ein formales Unendlichwerden der Amplitude der Gleichgewichtsbahn, auf für ganzzahliges $\tilde{Q}$, also an den Grenzen eines in der Nähe ganzzahligen Wertes von Q liegenden instabilen Bereichs. An den Grenzen jeden instabilen Bereichs gehen $\tilde{r}_1(\vartheta)$ und $\tilde{r}_2(\vartheta)$ in ein und dieselbe periodische oder antiperiodische Lösung $\tilde{r}_{\mathrm{per}}(\vartheta)$ über [24]; da dann die Wronskideterminante $\tilde{W}$ verschwindet, würde man zunächst Resonanz der geänderten Gleichgewichtsbahn an den Grenzen aller instabilen Bereiche erwarten. Eine genauere Untersuchung wird jedoch gleich zeigen, daß bei halbganzem Q keine Resonanz der Gleichgewichtsbahn auftritt. Die geänderte

[21] Vgl. Unterabschn. *a* der Einleitung, insbesondere Gl. (E.13). Daß die dort vorgetragenen Überlegungen sich zunächst auf die instabilen Bereiche für eine einzelne Periode des Feldindex bezogen, ist mathematisch ohne Bedeutung.

Gleichgewichtsbahn existiert mathematisch auch innerhalb der instabilen Bereiche; praktisch ist das allerdings ohne Interesse, da die Betatronschwingungen um diese Bahn herum dort nicht stabil sind.

Es soll das Verhalten der geänderten Gleichgewichtsbahn in der Umgebung eines instabilen Bereichs mit halbganzem Wert von $\tilde{Q}$ untersucht werden. In der Nähe einer solchen Bereichsgrenze kann näherungsweise gesetzt werden

$$1 - \exp\left[\pm 2\pi i \tilde{Q}\right] = 2 \,. \tag{I.3.3}$$

Daher folgt in diesem Fall aus Gl. (I.3.2)

$$\hat{r}(\vartheta) = \frac{1}{2\tilde{W}} \int_{\vartheta - 2\pi}^{\vartheta} \mathrm{d}\vartheta' \, \delta f(\vartheta') \{\tilde{r}_1(\vartheta')\tilde{r}_2(\vartheta) - \tilde{r}_2(\vartheta')\tilde{r}_1(\vartheta)\} \,. \tag{I.3.4}$$

Jetzt steht im Zähler ebenso wie im Nenner eine aus den Lösungen $\tilde{r}_1(\vartheta)$, $\tilde{r}_2(\vartheta)$ gebildete Determinante. Ein Quotient zweier solcher Determinanten ändert seinen Wert nicht, wenn zu irgend zwei anderen linear unabhängigen Lösungen übergegangen wird und die Tatsache, daß zwei spezielle Lösungen im Grenzfall ($\tilde{Q} \to$ nächste halbganze Zahl) gegen ein und dieselbe Funktion konvergieren, ist ohne Belang.

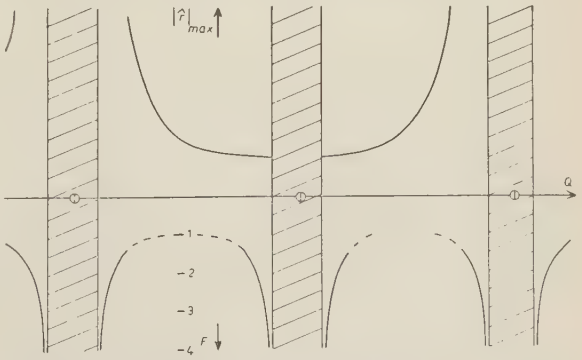


Abb. 6. – Schematische Darstellung der Auswirkung von Fehlern bei Änderung von Q infolge Änderung des idealen Feldindex: instabile Bereiche (die beiden äußeren für ganzzahliges Q), Amplitude der geänderten Gleichgewichtsbahn, Schwebungsfaktor.

In Abb. 6 ist das Zusammenwirken der Fehler qualitativ dargestellt. Zunächst sind die instabilen Bereiche eingezeichnet, die bei kleinen Fehlern des Feldindex in der Nähe ganz- oder halbzahliger Werte von Q liegen. Nach unten ist der Schwebungsfaktor F (Unterabschnitt 2*c*, insbes. Gl. (I.2.44)) aufgetragen, der an den Grenzen all dieser instabilen Bereiche gegen unendlich geht. Nach oben ist die Amplitude der geänderten Gleichgewichtsbahn eingezeichnet, die nur an den Grenzen der bei ganzzahligem Q liegenden instabilen Bereiche gegen unendlich geht, an den Grenzen der anderen Bereiche aber kein ungewöhnliches Verhalten zeigt. Die Betatronschwingungen mit ihren Schwebungen erfolgen um die geänderte Gleichgewichtsbahn herum. Der Abstand, den man von einem instabilen Bereich bei ganzzahligem Q einhalten muß, wird meist durch die maximal zulässige Auslenkung der geänderten Gleichgewichtsbahn bestimmt sein; bei den instabilen Bereichen mit halbganzem Q

ist der einzuhaltende Abstand durch den Schwebungseffekt und die damit verbundene Aufblähung des Strahls gegeben. Ferner wird zu beachten sein, daß auf Teilchen mit Impulsfehler ein etwas geänderter Feldindex wirkt; hierauf hat TWISS hingewiesen (siehe Anh. 1*a*).

Bei dem bisherigen qualitativen Überblick wurde von Näherungsausdrücken für $\tilde{r}_1(\vartheta)$, $\tilde{r}_2(\vartheta)$, $\tilde{Q}$ kein Gebrauch gemacht. Jetzt sollen bei der Untersuchung des Resonanzverhaltens die Ergebnisse der Störungsrechnung des vorangehenden Abschnitts benutzt werden; die folgenden Resultate sind daher nur für hinreichend kleine Fehler des Feldindex richtig. Setzt man in Gl. (I.3.1) für $\tilde{r}_1(\vartheta)$, $\tilde{r}_2(\vartheta)$ die Ausdrücke (I.2.34) ein, so folgt zunächst

$$\tilde{W}=(|a_1|^2-|a_2|^2)W\,. \tag{I.3.5}$$

Ferner kann im Nenner von Gl. (I.3.2) gesetzt werden

$$1-\exp[\pm 2\pi i\tilde{Q}]\approx\mp 2\pi i\tilde{Q}^{(1)}\,. \tag{I.3.6}$$

Im Zähler dagegen darf man meist setzen

$$\tilde{r}_1(\vartheta')\tilde{r}_2(\vartheta)\approx\tilde{r}_2(\vartheta')\tilde{r}_1(\vartheta)\approx\tilde{r}^*_{\mathrm{per}}(\vartheta')\tilde{r}_{\mathrm{per}}(\vartheta)\,, \tag{I.3.7}$$

wobei $\tilde{r}_{\mathrm{per}}(\vartheta)$ die periodische Lösung ($\tilde{r}_1(\vartheta)$ umittelbar auf der Bereichsgrenze) bedeutet. Nur falls das hier auftretende Integral extrem klein ist oder verschwindet, ist eine sorgfältigere Behandlung erforderlich; vgl. hierzu den Schluß dieses Unterabschnitts. Man erhält mit (I.3.5), (I.3.6) und (I.3.7) für die geänderte Gleichgewichtsbahn in der Nähe eines ganzzahligen Q näherungsweise

$$\hat{\tilde{r}}(\vartheta)=\frac{(r_1(\vartheta)+(a_2/a_1)r_2(\vartheta))\int_0^{2\pi}\delta f(\vartheta')\big((a_2^*/a_1^*)r_1(\vartheta')+r_2(\vartheta')\big)\,\mathrm{d}\vartheta'}{\pi i W(1-|a_2/a_1|^2)\tilde{Q}^{(1)}}\,. \tag{I.3.8}$$

Das Herausdividieren von $|a_1|^2$ ist erlaubt, obwohl diese Größe in Zähler und Nenner nicht genau identisch ist. In $\tilde{W}$ bezieht sie sich nämlich auf einen bestimmten Arbeitspunkt in der Nähe der Grenze des instabilen Bereichs, in $\tilde{r}_{\mathrm{per}}(\vartheta)$ dagegen auf diese Grenze selbst; die Unterschiede sind aber von höherer Ordnung.

Aus Gl. (I.2.41) folgt, daß der Betrag von a_2/a_1 an der Grenze des instabilen Bereichs gleich eins ist. Es kann also im Zähler der Gl. (I.3.8) gesetzt werden

$$\frac{a_2}{a_1}=\varepsilon\,, \tag{I.3.9}$$

wobei ε eine komplexe Zahl vom Betrag eins darstellt. Für den Nenner werden nun Gl. (I.2.41) und Gl. (I.2.43) benutzt. Somit erhält man schließlich für $\delta Q^{(1)}_{\text{abst}} \ll \delta Q^{(1)}_{\text{inst}}$

$$\hat{r}(\vartheta) = -\frac{(r_1(\vartheta) + \varepsilon r_2(\vartheta)) \int_0^{2\pi} \delta f(\vartheta')(\varepsilon^* r_1(\vartheta') + r_2(\vartheta')\,\mathrm{d}\vartheta'}{4\pi i W \delta Q^{(1)}_{\text{abst}}}, \tag{I.3.10}$$

wobei der Nenner also linear mit $\delta Q^{(1)}_{\text{abst}}$ geht. Die entstehende Gleichung ist sehr ähnlich zu dem Grenzfall der Gl. (I.1.8) (Gleichgewichtsbahn in einem Teilchenbeschleuniger ohne Fehler des Feldindex) in der Nähe eines ganzzahligen Wertes von Q

$$\hat{r}(\vartheta) = \frac{\int_0^{2\pi} \mathrm{d}\vartheta'\, \delta f(\vartheta')(r_1(\vartheta') r_2(\vartheta) + r_2(\vartheta') r_1(\vartheta))}{2\pi i W\, \delta Q}. \tag{I.3.11}$$

Der Betrag des Faktors von $r_1(\vartheta)$ ist ein Maß für die Amplitude der geänderten Gleichgewichtsbahn. Ohne Fehler im Feldindex ist dieser Faktor gegeben durch $(2\pi i W\, \delta Q)^{-1} \int_0^{2\pi} \delta f(\vartheta') r_2(\vartheta')\, \delta\vartheta'$, bei Anwesenheit von Fehlern dagegen durch $\left(2\pi i W\, \delta Q^{(1)}_{\text{abst}}\right)^{-1} \frac{1}{2} \int_0^{2\pi} \delta f(\vartheta')(r_2(\vartheta') + \varepsilon^* r_1(\vartheta'))\,\mathrm{d}\vartheta'$, wobei $\int_0^{2\pi} \delta f(\vartheta') r_1(\vartheta')\,\mathrm{d}\vartheta'$ und $\int_0^{2\pi} \delta f(\vartheta') \cdot r_2(\vartheta')\,\mathrm{d}\vartheta'$ konjugiert komplex sind. Je nach der Phase von ε hat man im zweiten Fall in einem bestimmten Punkt eine größere oder kleinere Amplitude als im ersten. Größenordnungsmäßig ergibt sich so, daß sich die Breite des gefährlichen Bereichs in der Umgebung einer Resonanz ungefähr additiv zusammensetzt aus der Breite des instabilen Bereichs und der Breite des Bereichs zu großer Auswanderung der Gleichgewichtsbahn, wie er sich allein aus den Fehlern des Führungsfeldes errechnen würde.

Falls für die Floquetschen Lösungen $r_1(\vartheta)$, $r_2(\vartheta)$ im idealen Beschleuniger an einer Resonanzstelle (Q ganzzahlig)

$$\int_0^{2\pi} \delta f(\vartheta') \begin{Bmatrix} r_1(\vartheta') \\ r_2(\vartheta') \end{Bmatrix} \mathrm{d}\vartheta' = 0 \tag{I.3.12}$$

gilt — etwa wegen strenger Periodizität von $\delta f(\vartheta)$ mit ϑ_0 —, so reicht die vereinfachte Gl. (I.3.10) nicht aus. Man hat dann unter dem Integral für $\tilde{r}_{\text{per}}(\vartheta)$ die erste störungstheoretische Näherung einzusetzen, d.h. nach Gl. (I.2.13) zu ersetzen

$$\int_0^{2\pi} \delta f(\vartheta) \tilde{r}_{\text{per}}(\vartheta)\,\mathrm{d}\vartheta \to \frac{1}{W} \int_0^{2\pi} \mathrm{d}\vartheta\, \delta f(\vartheta) \int_0^{\vartheta} \mathrm{d}\vartheta'\, \delta n(\vartheta') \tilde{r}_{\text{per}}(\vartheta') \{r_1(\vartheta') r_2(\vartheta) - r_2(\vartheta') r_1(\vartheta)\}, \tag{I.3.13}$$

wobei $\tilde{r}_{per}(\vartheta)$ auf der rechten Seite wieder durch den Ausdruck nullter Näherung darzustellen ist. Diese Überlegung gilt speziell für das Verhalten der synchronen Bahn unter dem Einfluß von Fehlern im Feldindex.

b) Ähnlichkeitsgesetze. – Zum Verständnis allgemeiner funktionaler Zusammenhänge, insbesondere solcher, die in Teil II auftreten werden, ist es zweckmäßig, den Begriff der «ähnlichen Anordnungen» einzuführen. Da ein Periodizitätsbereich des Feldindex in einem Synchrotron mit starker Stabilisierung — wenigstens, wenn man sich auf die Untersuchung einer einzigen Schwingungsrichtung beschränkt —, gekennzeichnet ist durch den Verlauf der Funktion $n(\vartheta)$ im Intervall 0 bis ϑ_0, sollen zwei Anordnungen als ähnlich bezeichnet werden, wenn die Feldindizes auseinander hervorgehen durch Multiplikation mit einem konstanten Faktor und Ähnlichkeitstransformation der ϑ-Achse. In Formeln: Es seien $n^{(1)}(\vartheta)$, $n^{(2)}(\vartheta)$ die Feldindizes der beiden Anordnungen; die Anordnungen sollen ähnlich genannt werden, wenn es zwei Konstanten λ und $\varkappa$ gibt, so daß

$$n^{(2)}(\vartheta) = \lambda n^{(1)}(\varkappa\vartheta) \tag{I.3.14}$$

für $0 \leqslant \vartheta < \vartheta_0^{(2)} = \vartheta_0^{(1)}/\varkappa$. In diesem Sinne sind also z.B. alle Anordungen zu einander ähnlich, bei denen jeder Periodizitätsbereich aus zwei Sektoren gleicher Länge mit einem Feldindex von gleichem Betrag aber entgegengesetztem Vorzeichen besteht (für solche Anordnungen galt Gl. (E.15)). Es sei betont, daß sich alle unsere Ähnlichkeitsüberlegungen auf fehlerfreie Anordnungen und nicht auf Fehler beziehen.

Die Sprechweise «zwei ähnliche Anordnungen führen zur gleichen Schwingungsform» soll bedeuten, daß sich zu jeder Lösung $r^{(1)}(\vartheta)$ der einen Anordnung eine andere $r^{(2)}(\vartheta)$ für die zweite Anordnung finden läßt, so daß gilt

$$r^{(2)}(\vartheta) = r^{(1)}(\varkappa\vartheta)\,. \tag{I.3.15}$$

Diese Bedingung bedeutet, daß die Betatronschwingungen in beiden Synchrotrons in gleicher Weise verlaufen. Aus der Differentialgleichung (E.6) für die freien Betatronschwingungen folgt, daß zwei ähnliche Anordnungen dann (und übrigens nur dann) zur gleichen Schwingungsform führen, wenn gilt

$$\varkappa = \sqrt{\lambda}\,. \tag{I.3.16}$$

Wegen der für den Periodizitätswinkel ϑ_0 gültigen Beziehung

$$\vartheta_0^{(2)} = \frac{\vartheta_0^{(1)}}{\varkappa} \tag{I.3.17}$$

kann Bedingung (I.3.16) auch so formuliert werden, daß die Größe $n^{\frac{1}{2}}\vartheta_0$ für alle zur gleichen Schwingungsform führenden ähnlichen Anordnungen den gleichen Wert hat. Das Zeichen n soll hier und auch später den Feldindex an einer bestimmten Stelle, also etwa bei $\vartheta = \alpha\vartheta_0$, bedeuten ($\alpha$ ist eine Konstante). Der Periodizitätswinkel ϑ_0 kann nach Gl. (E.8) durch die Periodenzahl M ausgedrückt werden. Damit gilt, daß ähnliche Anordnungen zur gleichen Schwingungsform führen, wenn die Größe $n^{\frac{1}{2}}/M$ für beide Anordnungen dieselbe ist. Insbesondere kann daher die Größe μ (Gl. (E.11)) für ähnliche Anordnungen nur von $n^{\frac{1}{2}}/M$ abhängen:

$$\mu = F\left(\frac{n^{\frac{1}{2}}}{M}\right). \tag{I.3.18}$$

Ein spezielles Beispiel stellt Gl. (E.15) dar.

Aus Gl. (I.3.18) folgt durch Differentiation

$$n\frac{\mathrm{d}\mu}{\mathrm{d}n} = -\frac{1}{2}\frac{n^{\frac{1}{2}}}{M}F'\left(\frac{n^{\frac{1}{2}}}{M}\right). \tag{I.3.19}$$

Da die rechte Seite wieder nur von $n^{\frac{1}{2}}/M$ abhängt, ist also der Quotient aus der Änderung von μ und der relativen Änderung des Feldindex für ähnliche Anordnungen gleicher Schwingungsform stets gleich. Für die Änderung der Zahl der Betatronwellenlängen auf dem Umfang folgt hiermit unter Benutzung von Gl. (E.12)

$$n\frac{\mathrm{d}Q}{\mathrm{d}n} = \frac{M}{2\pi}n\frac{\mathrm{d}\mu}{\mathrm{d}n} = -\frac{n^{\frac{1}{2}}}{4\pi}F'\left(\frac{n^{\frac{1}{2}}}{M}\right). \tag{I.3.20}$$

Die Dichte der Resonanzstellen, gemessen in relativen Änderungen des Feldindex, ist also umgekehrt proportional zu $n^{\frac{1}{2}}$.

Trifft man für die Lösungen $r^{(1)}(\vartheta)$, $r^{(2)}(\vartheta)$ die Zuordnung (I.3.15), so gilt ersichtlich für die Ableitungen

$$r^{(2)'}(\vartheta) = \varkappa r^{(1)'}(\varkappa\vartheta) \tag{I.3.21}$$

und entsprechend für die aus zwei Lösungen gebildete Wronskideterminante

$$W^{(2)} = \varkappa W^{(1)}. \tag{I.3.22}$$

Von mehr Interesse sind Bildungen, die von der Normierung der Lösungen unabhängig sind. Seien $r_1^{(1)}(\vartheta)$, $r_2^{(1)}(\vartheta)$, $r_1^{(2)}(\vartheta)$, $r_2^{(2)}(\vartheta)$ zwei entsprechende Paare von Lösungen und seien $W^{(1)}$, $W^{(2)}$ die aus diesen Lösungen gebildeten Wronski-

determinanten, so gilt ersichtlich

$$\frac{1}{W^{(2)}} r_1^{(2)}(\alpha\vartheta_0^{(2)}) r_2^{(2)}(\beta\vartheta_0^{(2)}) = \frac{1}{\varkappa W^{(1)}} r_1^{(1)}(\alpha\vartheta_0^{(1)}) r_2^{(1)}(\beta\vartheta_0^{(1)}) \tag{I.3.23}$$

$(0 \leqslant \alpha, \beta \leqslant 1)$. Die Größe $(n^{\frac{1}{2}}/W) r_1(\alpha\vartheta_0) r_2(\beta\vartheta_0)$ ist also invariant für ähnliche Anordnungen, die zur gleichen Schwingungsform führen; für ähnliche Anordnungen im allgemeinen hängt sie daher nur von $n^{\frac{1}{2}}/M$ ab. Für $r_1(\vartheta)$ und $r_2(\vartheta)$ kann man hier speziell die Floquetschen Lösungen einsetzen.

Aus der Transformation der Winkel ϑ, die in Gl. (I.3.17) speziell für den Periodizitätswinkel ϑ_0 angegeben wurde, folgt (in symbolischer Schreibweise) für Integrale

$$\int_0^{\vartheta_0^{(2)}} \mathrm{d}\vartheta^{(2)} \cdot (2) = \frac{1}{\varkappa} \int_0^{\vartheta_0^{(1)}} \mathrm{d}\vartheta^{(1)} \cdot (1) \,. \tag{I.3.24}$$

Die Beziehung bleibt richtig, wenn auf beiden Seiten über denselben Bruchteil des Periodizitätsbereichs integriert wird. Unter dem Integral der linken Seite sollen Größen der Anordnung 1, unter dem der rechten Seite die entsprechenden der Anordnung 2 stehen. Für die unter dem Integral stehenden Ausdrücke sind dabei Beziehungen vom Typ (I.3.23) zu benutzen. Symbolisch folgt aus Gl. (I.3.24), daß $n^{\frac{1}{2}} \int_0^{\vartheta_0} \mathrm{d}\vartheta \ldots$ für ähnliche Anordnungen nur von der die Schwingungsform kennzeichnenden Größe $n^{\frac{1}{2}}/M$ (bzw. von μ) abhängt.

Eine Reihe von Anwendungen dieser Überlegungen werden in Teil II auftreten. Hier soll nur eine Anwendung auf die synchrone Bahn (vgl. Unterabschn. 1*c*) erfolgen. Auf der rechten Seite der Gl. (I.1.17) stehen nämlich genau Integrale der eben betrachteten Art. Falls $\varrho(\vartheta)/R$ nur von der relativen Lage zwischen 0 und ϑ_0, für $\vartheta = \alpha\vartheta_0$ also nur von α abhängt, folgt somit, daß $r_{\text{synchr}}(\vartheta)/R$ für ähnliche Anordnungen gleicher Schwingungsform bei gleichem $\delta p/p$ umgekehrt proportional zu n ist. Ebenso folgt aus Gl. (I.1.19), daß der Quotient aus relativer Wegverlängerung und relativem Impulsfehler umgekehrt proportional zu n ist. Als Spezialfall vergleiche man etwa die in CLS, Gln. (8) und (9) angegebenen Beziehungen.

4. – Zusammenfassung von Teil I.

Die mathematische Analyse hat die qualitativen Überlegungen des Unterabschnitts *b* der Einleitung bestätigt. Es wurden zwei Typen von Abweichungen eines Teilchenbeschleunigers mit starker Stabilisierung vom idealen Entwurf betrachtet, nämlich Abweichungen im magnetischen Führungsfeld auf der

Gleichgewichtsbahn (Abschn. 1) und im Feldindex (Abschn. 2). Der dritte Typ linearer Fehler, nämlich eine Verwindung der Magnete gegeneinander, wurde hier nicht behandelt.

Ein ideales Synchrotron mit starker Stabilisierung besteht aus einer Anzahl streng identischer Periodizitätsbereiche des Feldindex mit einer in sich zurücklaufenden Gleichgewichtsbahn. Für einen solchen idealen Beschleuniger gilt die Theorie der starken Stabilisierung in ihrer ursprünglichen Formulierung (Unterabschn. *a* der Einleitung). Treten nur Fehler im Führungsfeld, jedoch nicht solche im Feldindex auf, so gibt es in dem Beschleuniger eine abgeänderte Gleichgewichtsbahn, deren Abweichung von der ursprünglichen Gleichgewichtsbahn umso größer wird, je mehr sich für den zugeordneten idealen Beschleuniger die Zahl der Betatronwellenlängen auf dem Umfang einer ganzen Zahl nähert. Die Gleichung dieser geänderten Gleichgewichtsbahn ist durch Gl. (I.1.8), mit $\delta f(\vartheta)$ nach Unterabschn. *b* der Einleitung, gegeben. Fehler im Feldindex führen zu instabilen Bereichen in der Nähe solcher Arbeitspunkte, die im idealen Beschleuniger eine ganze oder halbganze Zahl von Betatronwellenlängen auf dem Umfang liefern. Die für die Bereiche kennzeichnenden Größen (ihre Lage und Breite) sind in erster störungstheoretischer Näherung in Gln. (I.2.30) und (I.2.31) angegeben. In der Nähe solcher instabilen Bereiche zeigen die Betatronschwingungen Schwebungen, ausgedrückt durch den Schwebungsfaktor F in Gl. (I.2.44) .Gleichzeitige Anwesenheit von Fehlern im Führungsfeld und im Feldindex führt zu einem Unendlichwerden der Amplitude der geänderten Gleichgewichtsbahn an den Grenzen derjenigen instabilen Bereiche, bei denen die Zahl der Betatronwellenlängen auf dem Umfang eine ganze Zahl wird (Abschn. 3*a*). Instabile Bereiche, Schwebungsfaktor und geänderte Gleichgewichtsbahn sind in Abb. 6 qualitativ dargestellt.

II. – Statistische Auswertung.

Die in Teil I abgeleiteten Gleichungen erlauben, die Auswirkung bekannter Fehler im Führungsfeld und im Feldindex auf die Betatronschwingungen zu berechnen. Die Fragestellung, die bei der Konstruktion eines Synchrotrons mit starker Stabilisierung auftritt, ist jedoch eine andere: es interessiert, Auswirkungen welcher Größe (Auslenkung der geänderten Gleichgewichtsbahn, Breite der instabilen Bereiche usw.) man bei dem noch nicht gebauten Beschleuniger zu erwarten hat. Es ergibt sich also ein statistisches Problem; aus begründeten Annahmen über die Größenverteilung der einzelnen Beiträge zu den Fehlern und ihrer gegenseitigen statistischen Abhängigkeit sind statistische Aussagen über die Größe der auftretenden Störungen zu gewinnen.

Teil II ist den dabei auftretenden Problemen gewidmet. In Abschn. 1 werden Mittelwerte (oder Erwartungswerte) für die interessierenden Größen

berechnet. Ein wirklicher Teilchenbeschleuniger ist jedoch weder mit dem idealen Entwurf identisch — das eben ist die Auswirkung der betrachteten Störungen —, noch besitzt er genau « mittlere » Fehler. Die auftretenden Auswirkungen der Fehler können deshalb sowohl größer als kleiner als die in Abschn. 1 zu berechnenden Mittelwerte sein. Darum erfolgt in Abschn. 2 eine Untersuchung der Wahrscheinlichkeitsverteilungen.

1. – Mittelwerte.

a) *Geänderte Gleichgewichtsbahn, Vorbemerkungen.* – Der statistische Charakter der Fragestellung soll zunächst am Beispiel der Fehler im magnetischen Führungsfeld genauer verdeutlicht werden ([25]). Versucht man, denselben Entwurf eines Synchrotrons in einer großen Zahl von Exemplaren zu verwirklichen, so werden die Fehler, z.B. Aufstellungsfehler, der einzelnen Verwirklichungen im allgemeinen verschieden sein. Auf diese Weise kann man eine Gesamtheit von Teilchenbeschleunigern erzeugt denken, die sämtlich Verwirklichungen desselben idealen Entwurfs darstellen. Ein Mitglied dieser Gesamtheit ist der Beschleuniger, den man bauen wird; man weiß jedoch nicht, welches. Ein gewisses Maß für die zu erwartenden Störungen wird man bekommen, indem man aus dieser Gesamtheit einen « mittleren » Beschleuniger herausgreift. Genauere Aufschlüsse erhält man, wenn man die Wahrscheinlichkeitsverteilung für die Größe der Fehlerauswirkungen aufstellt.

Vor Konstruktion der Gesamtheit, über die dann die Mittelbildung auszuführen ist, müssen die statistischen Eigenschaften der einzelnen Fehler untersucht werden. Es seien etwa die radialen Aufstellungsfehler betrachtet. Die Aufstellung kann beispielsweise so erfolgen, daß Abstand und Richtung jedes einzelnen Magneten in Bezug auf den Mittelpunkt des Beschleunigers festgelegt werden, wobei die Vermessung natürlich nur mit einer endlichen Genauigkeit und einer bestimmten Fehlerverteilung vorgenommen werden kann. In der Gesamtheit von Beschleunigern sind dann, wenigstens soweit es sich um diesen speziellen Fehler handelt, die Aufstellungsfehler der einzelnen Magnete von einander statistisch unabhängig und die relative Häufigkeit der einzelnen Fehler ist durch die Fehlerverteilung des Vermessungsprozesses gegeben. Nimmt man die Justierung der einzelnen Magnete nicht relativ zum Mittelpunkt des Beschleunigers sondern teilweise relativ zu den Nachbarn oder zu einzelnen Hilfspunkten vor, so sind die Aufstellungsfehler der einzelnen Magnete nicht mehr innerhalb der Gesamtheit statistisch unabhängig von einander. Das wirkt sich

---

([25]) Unseres Wissens war LAWSON (vgl. Fußn. ([11])) der erste, der eine statistische Aussage über den Einfluß von Fehlern auf den Betrieb des Synchrotrons mit starker Stabilisierung machte.

auf die Berechnung von Mittelwerten und Wahrscheinlichkeitsverteilungen natürlich aus.

Bei gegebenen Fehlern im magnetischen Führungsfeld, ausgedrückt durch die Funktion $\delta f(\vartheta)$ (vgl. Unterabschn. *b* der Einleitung), läßt sich die Abweichung $\hat{r}(\vartheta)$ der geänderten Gleichgewichtsbahn von der idealen Gleichgewichtsbahn mittels Gl. (I.1.8) berechnen. Da der Teilchenstrahl während des ganzen Beschleunigungszyklus innerhalb der Vakuumkammer bleiben muß, interessiert die maximale Auslenkung dieser geänderten Gleichgewichtsbahn. Diese Auslenkung muß stets kleiner bleiben als der Halbmesser des Vakuumrohres, da um die Gleichgewichtsbahn herum die freien Betatronschwingungen erfolgen. Bei der horizontalen Schwingungsrichtung ist außerdem zu beachten, daß die Synchrotronschwingungen (vgl. Teil I, Abschn. 1*c*) zu einer zusätzlichen horizontalen Auslenkung führen.

Für die statistische Untersuchung der Auslenkung der geänderten Gleichgewichtsbahn wäre eigentlich folgendes Verfahren wünschenswert: Für jedes Mitglied der Gesamtheit, die entsprechend der statistischen Eigenschaften der einzelnen Fehler konstruiert ist, berechnet man nach Gl. (I.1.8) die maximale Auslenkung der Gleichgewichtsbahn. Anschließend ermittelt man hieraus den Mittelwert dieser Maximalauslenkung — gemittelt über die Gesamtheit von Beschleunigern —, die Wahrscheinlichkeitsverteilung dieser Größe usw. Die Durchführung dieses Programms würde recht schwierig sein, im wesentlichen deswegen, weil sich der Gl. (I.1.8) nicht auf einfache Weise Lage und Wert des Maximums ansehen lassen. Es empfiehlt sich deshalb ein anderes, im Einzelfall weniger genaues Verfahren, das jedoch etwa dieselben Aufschlüsse über die zu erwartenden Auswirkungen der Fehler gewinnen läßt.

Zunächst ist nach Gl. (I.1.8) sicherlich die folgende Ungleichung zutreffend

$$|\hat{r}|_{\max} \leqslant \frac{|r_1|_{\max} \cdot |\int \delta f(\vartheta') r_1(\vartheta') \,\mathrm{d}\vartheta'|_{\max}}{|W| \cdot |\sin \pi Q|}, \tag{II.1.1}$$

wobei benutzt wurde

$$|1 - \exp[-2\pi i Q]|^2 = 2(1 - \cos 2\pi Q) = 4 \sin^2 \pi Q . \tag{II.1.2}$$

In Gl. (II.1.1) bedeutet $|r_1|_{\max}$ den Maximalbetrag der Funktion $r_1(\vartheta)$, der in jedem Periodizitätsbereich des Feldindex (mindestens) einmal angenommen wird. In Anordnungen, bei denen jede Periode aus zwei Sektoren konstanten Feldindexes besteht, wird dieses Maximum aus Symmetriegründen (für die horizontale Schwingungsrichtung) in der Mitte des Sektors mit negativem Feldindex angenommen. $|\int \delta f(\vartheta') r_1(\vartheta') \,\mathrm{d}\vartheta'|_{\max}$ bedeutet das Maximum von $|\int_{\vartheta-2\pi}^{\vartheta} \delta f(\vartheta') r_1(\vartheta') \,\mathrm{d}\vartheta'|$, das jeweils für ein gewisses, von Beschleuniger zu Be-

schleuniger verschiedenes, ϑ angenommen wird. Durch Gl. (II.1.1) wird die Untersuchung des Maximums von $|\hat{r}|$ auf die des Maximums von $|\int_{\vartheta-2\pi}^{\vartheta}\delta f(\vartheta')r_1(\vartheta')\mathrm{d}\vartheta'|$ zurückgeführt. Die Maximalamplitude wird in jedem einzelnen Fall damit möglicherweise überschätzt, aber nie unterschätzt.

Die Untersuchung von $|\int\delta f(\vartheta')r_1(\vartheta')\mathrm{d}\vartheta'|$ soll in zwei Teilen erfolgen. Zunächst (Unterabschn. *b*) werde diese Größe an einem beliebig herausgegriffenen Punkt der Bahn berechnet und dann eine geeignete Mittelung über die Gesamtheit von Beschleunigern durchgeführt. Durch diese verhältnismäßig einfachen Überlegungen wird bereits eine Aussage über die Auslenkung der geänderten Gleichgewichtsbahn gewonnen, die umso zuverlässiger ist, je mehr man in der Nähe von Resonanzbedingungen (Q ganzzahlig) arbeitet. Das Integral $\int_{\vartheta-2\pi}^{\vartheta}\delta f(\vartheta')r_1(\vartheta')\mathrm{d}\vartheta'$ ist dann nämlich von dem Winkel ϑ praktisch unabhängig, da die Funktion $r_1(\vartheta)$ nahezu periodisch mit der Periode 2π ist. Man hat in der Nähe kritischer Bedingungen als geänderte Gleichgewichtsbahn praktisch eine in sich zurücklaufende Betatronschwingung. Die Schwankung von $|\int_{\vartheta-2\pi}^{\vartheta}\delta f(\vartheta')r_1(\vartheta')\mathrm{d}\vartheta'|$ entlang des Umfangs spielt eine größere Rolle, falls man sich ziemlich weit von Resonanzbedingungen entfernt befindet. Infolge dieser Schwankungen kann dann das Maximum von $|\int_{\vartheta-2\pi}^{\vartheta}\delta f(\vartheta')r_1(\vartheta')\mathrm{d}\vartheta'|$ merklich größer sein als der Wert in einem beliebig herausgegriffenen Punkt. Eine Untersuchung der Schwankungen der genannten Größe entlang des Umfangs des Beschleunigers erfolgt in Unterabschn. *c*.

b) Geänderte Gleichgewichtsbahn, Vernachlässigung der Amplitudenschwankung entlang des Umfangs. – Nach den Überlegungen von Unterabschn. *a* ist die Größe $\int_{\vartheta-2\pi}^{\vartheta}\delta f(\vartheta')r_1(\vartheta')\mathrm{d}\vartheta'$ für einen festen Punkt auf dem Umfang, etwa $\vartheta=2\pi$, zu berechnen. Es werde hierbei die Abkürzung ([26])

$$F=\int_0^{2\pi}\delta f(\vartheta)r_1(\vartheta)\mathrm{d}\vartheta \tag{II.1.3}$$

([26]) Eine Verwechslung dieser Größe F mit dem Schwebungsfaktor in Teil I, Abschn. 2*c* ist wohl nicht zu befürchten.

eingeführt. Ferner *definiere* man einen Mittelwert $\langle \hat{r}_{\max} \rangle$ durch ([27,28])

$$\langle \hat{r}_{\max} \rangle = \frac{|r_1|_{\max} \sqrt{\langle |F|^2 \rangle}}{|W|\,|\sin \pi Q|} \,. \tag{II.1.4}$$

Gl. (II.1.4) würde eine bessere Abschätzung für die Auslenkung der geänderten Gleichgewichtsbahn geben, wenn man hier statt F an der willkürlichen Stelle $\vartheta = 2\pi$ den jeweiligen Maximalwert einsetzen würde. Nunmehr werde Gebrauch gemacht von den Periodizitätseigenschaften (E.10) der Funktion $r_1(\vartheta)$ und gesetzt

$$F = \sum_{m=1}^{M} \varphi_m \exp\left[i(m-1)\mu\right] , \tag{II.1.5}$$

wobei

$$\varphi_m = \int_{(m-1)\vartheta_0}^{m\vartheta_0} \delta f(\vartheta)\left(r_1(\vartheta) \exp\left[-i(m-1)\mu\right]\right) \mathrm{d}\vartheta \tag{II.1.6}$$

ein Integral über den m. Periodizitätsbereich des Feldindex bedeutet. Die in Gl. (II.1.6) unter dem Integralzeichen stehende Funktion $r_1(\vartheta) \exp\left[-i(m-1)\mu\right]$ ist so gewählt, daß sie von der Nummer (« m ») des betreffenden Periodizitätsbereichs unabhängig ist.

Es werde nun angenommen, daß positive und negative Fehler $\delta f(\vartheta)$ gleich wahrscheinlich sind, daß also gilt

$$\langle \varphi_m \rangle = \langle \varphi_m^* \rangle = 0 \,. \tag{II.1.7}$$

Außerdem sollen diese Fehler in allen Periodizitätsbereichen in gleicher Weise auftreten und Fehler in verschiedenen Periodizitätsbereichen statistisch unabhängig von einander sein; man hat dann

$$\langle \varphi_m \varphi_l^* \rangle = \delta_{ml} \langle |\varphi|^2 \rangle \tag{II.1.8}$$

([27]) Eine Verbesserung dieser Definition (d.h. eine bessere Anpassung an das wirkliche Problem) wird nach Unterabschn. *c* erreicht, indem man $1/|\sin \pi Q|$ ersetzt durch $1/|\sin \pi Q| + \varepsilon$, wobei ε eine Zahl der Größenordnung eins darstellt.

([28]) Mittelwerte über die gedachte Gesamtheit von Beschleunigern sollen, wie es in der englischsprachigen Literatur üblich ist, durch spitze Klammern bezeichnet werden. Das ebenfalls gebräuchliche Zeichen eines Querstrichs über der betreffenden Größe soll der Mittelbildung entlang des Umfangs eines einzelnen Beschleunigers vorbehalten werden (vgl. Gl. (II. 1.23)).

(δ_{ml} ist das Kroneckersymbol). Aus Gln. (II.1.5), (II.1.7), (II.1.8), erhält man

$$\langle|F|^2\rangle = M\langle|\varphi|^2\rangle\,. \tag{II.1.9}$$

Mit der Definition (II.1.4) gewinnt man so schließlich

$$\langle\hat{r}_{\max}\rangle = \frac{\sqrt{M}\,|r_1|_{\max}\langle|\varphi|\rangle_{WmQ}}{|W|\cdot|\sin\pi Q|}\,. \tag{II.1.10}$$

Dabei wurde die Abkürzung

$$\langle|\varphi|\rangle_{WmQ} = \sqrt{\langle|\varphi|^2\rangle} \tag{II.1.11}$$

(WmQ = Wurzel aus dem mittleren Quadrat) benutzt. Man erkennt in Gl. (II.1.10) den Resonanznenner, der jeweils bei ganzzahligem Wert von Q verschwindet. Das Auftreten von $\sqrt{M}$ ist charakteristisch für die Annahme der statistischen Unabhängigkeit der Fehler in den einzelnen Periodizitätsbereichen.

Die genauere Auswertung der Gl. (II.1.10) sei zunächst für Aufstellungsfehler vorgenommen; die Voraussetzung statistischer Unabhängigkeit der Fehler in den einzelnen Periodizitätsbereichen ist erfüllt, wenn diese Bereiche unabhängig von einander relativ zum Mittelpunkt des Beschleunigers justiert werden. Es sei darüber hinaus jetzt etwa angenommen, daß jeder Periodizitätsbereich aus zwei Sektoren besteht, die einzeln relativ zum Mittelpunkt des Beschleunigers vermessen sind. Es kann sich hierbei beispielsweise um die beiden Stücke konstanten Feldindex in einer zu Gl. (E.15) führenden Anordnung handeln; von der Voraussetzung eines konstanten Feldindex wird jedoch im folgenden kein Gebrauch gemacht. Die Größe φ_m (Gl. (II.1.6)) spaltet auf in zwei Anteile

$$\begin{cases} \varphi_m = \varphi_m^{(1)} + \varphi_m^{(2)}\,, \\ \varphi_m^{(1)} = \displaystyle\int_{\substack{\text{1. Sektor des}\\ \text{m. Period.-Ber.}}} \delta f(\vartheta)\, r_1(\vartheta)\,\mathrm{d}\vartheta\,, \qquad \varphi_m^{(2)} = \int_{\substack{\text{2. Sektor des}\\ \text{m. Period.-Ber.}}} \delta f(\vartheta)\, r_1(\vartheta)\,\mathrm{d}\vartheta\,, \end{cases} \tag{II.1.12}$$

und man hat

$$\langle|\varphi|^2\rangle = \langle|\varphi^{(1)}|^2\rangle + \langle|\varphi^{(2)}|^2\rangle\,. \tag{II.1.13}$$

Der horizontale Aufstellungsfehler jedes einzelnen Sektors besteht aus einem Fehler $\delta\xi$ in der Aufstellung des Mittelpunktes und aus einem Winkelfehler $\delta\eta$; für die Größe $\delta\xi(\vartheta)$ in Gl. (E.16) hat man somit

$$\delta\xi(\vartheta) = \delta\xi + \delta\eta\cdot R(\vartheta - \vartheta_m)\,, \tag{II.1.14}$$

wobei diese Gleichung innerhalb jedes einzelnen Sektors gelten und ϑ_m die zum Mittelpunkt des Sektors gehörige Koordinate ϑ bedeuten soll. Nimmt man die Fehler $\delta\xi$ und $\delta\eta$ als unabhängig von einander, ihre Verteilungen für beide Sektoren aber als gleich an, so ergibt sich schließlich [29]

$$\text{(II.1.15)} \qquad \langle \hat{r}_{\max} \rangle = \frac{\sqrt{M}\,|r_1|_{\max}}{|\sin \pi Q|\,W} \left\{ \langle \delta\xi^2 \rangle \left(\left| \int_{1.\,\text{Sektor}} n(\vartheta)\, r_1(\vartheta)\, \mathrm{d}\vartheta \right|^2 + \left| \int_{2.\,\text{Sektor}} n(\vartheta)\, r_1(\vartheta)\, \mathrm{d}\vartheta \right|^2 \right) + \right.$$

$$\left. + \langle \delta\eta^2 \rangle \left(\left| \int_{1.\,\text{Sektor}} n(\vartheta)(\vartheta - \vartheta_m)\, r_1(\vartheta)\, \mathrm{d}\vartheta \right|^2 + \left| \int_{2.\,\text{Sektor}} n(\vartheta)(\vartheta - \vartheta_m)\, r_1(\vartheta)\, \mathrm{d}\vartheta \right|^2 \right) \right\}^{\frac{1}{2}}.$$

Bei einer feineren Unterteilung der Perioden des Feldindex in unabhängig justierte Sektoren steht in der geschweiften Klammer eine entsprechend größere Anzahl von Termen.

Auf die hier auftretenden Integrale können die Ähnlichkeitsgesetze aus Teil I, Abschn. 3*b* angewandt werden. Man findet, daß der Ausdruck $r_1|_{\max} |W|^{-1} \int n(\vartheta) r_1(\vartheta)\, \mathrm{d}\vartheta$ (mit Integration über einen Sektor) für ähnliche Anordnungen nur von $n^{\frac{1}{2}}/M$ abhängt, der Ausdruck $|r_{1,\max}|\, W|^{-1} \int n(\vartheta)(\vartheta - \vartheta_m) \cdot r_1(\vartheta)\, \mathrm{d}\vartheta$ dagegen bei gegebenem $n^{\frac{1}{2}}/M$ proportional zu $n^{\frac{3}{2}}$ ist. Wenn man Winkelfehler für den Augenblick vernachlässigt, so ergibt sich also bei festem $n^{\frac{1}{2}}/M$ der Mittelwert der Amplitude der geänderten Gleichwichtsbahn — genaugenommen im Sinne der Definition (II.1.14) — proportional zu $\sqrt{M}$, d.h. (bei festem $n^{\frac{1}{2}}/M$) zu $n^{\frac{1}{4}}$. Dieselbe Abhängigkeit gilt für den Winkelfehler, falls nicht unmittelbar eine Winkelmessung sondern eine Messung der Koordinaten der Sektorenendpunkte erfolgt; für höheres n werden die Sektoren kürzer und dadurch wird der oben genannte Faktor $n^{-\frac{1}{2}}$ gerade kompensiert. Mit zunehmendem Feldindex nimmt demnach nicht nur die Dichte der Resonanzpunkte (proportional zu $n^{\frac{1}{2}}$, vgl. Gl. (I.3.20)) zu, auch die Amplitude der geänderten Gleichgewichtsbahn wird wegen der größeren Zahl unabhängig zu justierender Elemente größer.

Nach diesen Überlegungen kann Gl. (II.1.15) für ähnliche Anordnungen folgendermaßen geschrieben werden [29]

$$\text{(II.1.16)} \qquad \hat{r}_{\max} \sim \frac{n^{\frac{1}{2}}}{\sin \pi Q} \left\{ \langle \delta\xi^2 \rangle \left(f_{\delta\xi}\left(\frac{n^{\frac{1}{2}}}{M} \right) \right)^2 + \langle \delta\eta^2 \rangle \frac{R^2}{n} \left(f_{\delta\eta}\left(\frac{n^{\frac{1}{2}}}{M} \right) \right)^2 \right\}^{\frac{1}{2}}.$$

[29] In Unterabschn. *c* wird gezeigt werden, daß sich die Schwankung von $\int_{\vartheta-2\pi}^{\vartheta} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta'$ entlang des Umfangs näherungsweise berücksichtigen läßt, indem man hier $1/|\sin \pi Q|$ ersetzt durch $1/|\sin \pi Q| + \varepsilon$, wobei ε eine positive Zahl der Größenordnung eins darstellt.

Die Funktionen

$$\text{(II.1.17)}\quad \begin{cases} f_{\delta\xi}\left(\frac{n^{\frac{1}{2}}}{M}\right) = \frac{|r_1|_{\max}}{|W|}\left(\frac{n^{\frac{1}{2}}}{M}\right)^{-\frac{1}{2}}\left(\left|\int\limits_{\text{1. Sektor}} n(\vartheta)\, r_1(\vartheta)\,\mathrm{d}\vartheta\right|^2 + \left|\qquad\qquad\right|^2_{\text{2. Sektor}}\right)^{\frac{1}{2}} \\ f_{\delta\eta}\left(\frac{n^{\frac{1}{2}}}{M}\right) = \frac{|r_1|_{\max}}{|W|}\left(\frac{n^{\frac{1}{2}}}{M}\right)^{-\frac{1}{2}}\left(\left|\int\limits_{\text{1. Sektor}} n(\vartheta)(\vartheta-\vartheta_m)\, r_1(\vartheta)\,\mathrm{d}\vartheta\right|^2 + \left|\qquad\qquad\right|^2_{\text{2. Sektor}}\right)^{\frac{1}{2}} \end{cases}$$

hängen dabei nur noch von der kennzeichnenden Größe $n^{\frac{1}{2}}/M$ und damit von μ ab. Zum Verständnis des bei dem Winkelfehler auftretenden Terms beachte man noch, daß gilt

$$\text{(II.1.18)}\quad \frac{R^2}{n} = \left(\frac{n^{\frac{1}{2}}}{M}\right)^{-2}\left(\frac{L}{2\pi}\right)^2$$

mit L als geometrischer Länge eines Periodizitätsbereichs. Da der Radius R jedoch durch die gewünschte Maximalenergie vorgegeben ist, ist eine Ersetzung von R durch L meist nicht zweckmäßig.

Die in Gl. (II.1.17) auftretenden Integrale können unter Benutzung der Bewegungsgleichung (E.6) noch weiter umgeformt werden. Man hat nämlich [30]

$$\text{(II.1.19)}\quad \begin{cases} \int n(\vartheta)\, r_1(\vartheta)\,\mathrm{d}\vartheta = [r_1'(\vartheta)]\,, \\ \int n(\vartheta)(\vartheta-\vartheta_m)\, r_1(\vartheta)\,\mathrm{d}\vartheta = [(\vartheta-\vartheta_m)\, r_1'(\vartheta)] - [r_1(\vartheta)]\,, \end{cases}$$

wobei rechts als Grenzen die beiden Enden des jeweiligen Sektors einzusetzen sind. Als Beispiel wurden die Funktionen $f_{\delta\xi}(n^{\frac{1}{2}}/M)$, $f_{\delta\eta}(n^{\frac{1}{2}}/M)$ für eine Anordnung mit Sektoren gleicher Länge und einem Feldindex gleichen Betrages aber entgegengesetzten Vorzeichens (vgl. Gl. (E.15)) berechnet; sie sind in Abb. 7 dargestellt. Es ist nicht verwunderlich, daß beide Funktionen an der oberen Grenze des Stabilitätsbereichs gegen unendlich gehen. Mathematisch folgt das aus dem Auftreten der Wronskideterminante im Nenner der die

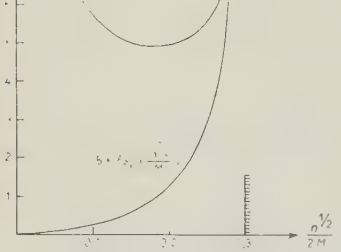


Abb. 7. – Die Funktionen $f_{\delta\xi}(n^{\frac{1}{2}}/M)$ und $f_{\delta\eta}(n^{\frac{1}{2}}/M)$, die den Einfluß statistischer Aufstellungsfehler (Lage $\delta\xi$ und Winkel $\delta\eta$) der einzelnen Sektoren auf die Auslenkung der geänderten Gleichgewichtsbahn beschreiben; vgl. (II.1.16).

[30] Eine eckige Klammer soll hier und in späteren Gleichungen bedeuten, daß die Differenz der Werte der Funktion an den beiden Integrationsgrenzen — meist den Endpunkten eines Sektors — zu bilden ist.

Funktionen definierenden Ausdrücke; die Wronskideterminante verschwindet nämlich an der Grenze eines Stabilitätsbereichs, da $r_1(\vartheta)$ und $r_2(\vartheta)$ dort gegen die gleiche periodische oder antiperiodische Funktion konvergieren. Bei $n^{\frac{1}{2}}/M = 0$ liegt im mathematischen Sinne keine Grenze des Stabilitätsbereichs vor; das Unendlichwerden von $f_{\delta\xi}(n^{\frac{1}{2}}/M)$ und einigen anderen später vorkommenden Funktionen rührt dort von der Abspaltung des Faktors $(n^{\frac{1}{2}}/M)^{\frac{1}{2}}$ her.

Wie man aus den berechneten Funktionen $f_{\delta\xi}(n^{\frac{1}{2}}/M)$ und $f_{\delta\eta}(n^{\frac{1}{2}}/M)$ sieht, spielen bei dieser Anordnung Winkelfehler im allgemeinen gar keine Rolle verglichen mit Fehlern in der Lage des Sektorenmittelpunkts. Diese Behauptung ist jedenfalls richtig, solange statt einer eigentlichen Winkelmessung eine Justierung der beiden Endpunkte eines Sektors vorgenommen wird. Es leuchtet auch anschaulich ein, daß sich die Störungen durch derartige Winkelfehler nahezu kompensieren, falls die Wellenlänge der Betatronschwingungen groß gegen die Sektorenlänge ist. Koppelt man dagegen stets einen halben Sektor mit positivem Feldindex und einen halben mit negativem Feldindex zusammen und justiert jetzt diese neuen Einheiten, so liegen die Verhältnisse gerade umgekehrt. Man hat dann relative Unempfindlichkeit gegen Fehler in der Lage der Mitte der Einheit, dafür aber eine entsprechend stärkere Empfindlichkeit gegen Winkelfehler.

In entsprechender Weise ist eine statistische Behandlung anderer Verfahren zur Justierung der Magnete möglich; es können etwa eine begrenzte Zahl von Punkten auf dem Umfang in Bezug auf den Mittelpunkt des Beschleunigers vermessen und die Magnete selbst dann relativ zu diesen Hilfspunkten justiert werden. Wesentlich neue grundsätzliche Gesichtspunkte ergeben sich hierbei nicht; deshalb soll auf die Einzelheiten nicht eingegangen werden.

Aufstellungsfehler sind nicht die einzigen Ursachen von Fehlern im magnetischen Führungsfeld; z.B. wirkt die statistische Schwankung der remanenten Felder in der gleichen Weise. Vermutlich können die Schwankungen $\delta H(\vartheta)$ der Feldstärken der einzelnen Sektoren als unabhängig von einander angenommen werden. Es ist jedoch schwierig zu sagen, ob die Feldfehler eines einzelnen Sektors durch ein über die ganze Sektorlänge konstantes δH hinreichend gut beschrieben werden. Für sehr lange Sektoren ist das wahrscheinlich nicht mehr richtig, für kürzere Sektoren wird diese Annahme vermutlich ausreichen. Mit dieser Annahme erhält man für die durch die Schwankung der remanenten Felder von Sektor zu Sektor hervorgerufene Amplitude $\langle \hat{r}_{\max} \rangle$ der geänderten Gleichgewichtsbahn — definiert durch Gl. (II.1.4) — folgenden Ausdruck [29]

$$\langle r_{\max} \rangle = \frac{n^{-\frac{3}{4}}}{|\sin \pi Q|} R^2 \left\langle \frac{\delta H}{H_\varrho} \right\rangle f_{\delta H}\left(\frac{n^{\frac{1}{2}}}{M}\right), \tag{II.1.20}$$

wobei die Funktion

$$f_{\delta H}\left(\frac{n^{\frac{1}{2}}}{M}\right) = \frac{|r_1|_{\max}}{|W|} n \left(\frac{n^{\frac{1}{2}}}{M}\right)^{-\frac{1}{4}} \left(\left| \int_{1.\,\text{Sektor}} r_1(\vartheta)\, \mathrm{d}\vartheta \right|^2 + \left| \int_{2.\,\text{Sektor}} \right|^2 \right)^{\frac{1}{2}} \tag{II.1.21}$$

nur von $n^{\frac{1}{2}}/M$ abhängt. Man sieht aus Gl. (II.1.20), daß diese Störung mit wachsendem Feldindex abnimmt. Hierin drückt sich die starke Stabilisierung aus. Bei Aufstellungsfehlern trat ein entsprechender Faktor n^{-1} nicht auf, da eine gegebene Verrückung bei größerem Feldindex zu einer entsprechend größeren Änderung des magnetischen Feldes führt. Speziell für eine Anordnung mit Sektoren konstanten Feldindexes gilt die einfache Beziehung

$$f_{\delta R}\left(\frac{n^{\frac{1}{2}}}{M}\right) = f_{\delta \xi}\left(\frac{n^{\frac{1}{2}}}{M}\right). \tag{II.1.22}$$

Treten mehrere unabhängige Fehler im Führungsfeld auf (z.B. Aufstellungsfehler und eigentliche Feldfehler), so ist das zugehörige $\langle r_{\max}\rangle$ durch die Quadratwurzel aus der Summe der Quadrate der einzelnen Anteile gegeben.

c) Geänderte Gleichgewichtsbahn, Amplitudenschwankung entlang des Umfangs. – Im vorangehenden Unterabschnitt wurde die Schwankung des Ausdrucks $\left|\int_{\vartheta-2\pi}^{\vartheta} \delta f(\vartheta') r_1(\vartheta')\,\mathrm{d}\vartheta'\right|$ entlang des Umfangs, also für verschiedene Werte von ϑ, vernachlässigt, indem der Quadratmittelwert dieser Größe für $\vartheta = 2\pi$ gewählt und hiermit die Amplitude der Gleichgewichtsbahn abgeschätzt wurde. Den damit gewonnenen Mittelwert kann man auch so verstehen, daß zunächst eine Mittelbildung über den Umfang eines einzelnen Beschleunigers (genauer: die Punkte $\vartheta = 0,\ \vartheta_0,\ 2\vartheta_0$, usw.)

$$\overline{|F|^2} = \frac{1}{M}\left\{\left|\int_{-2\pi}^{0} \delta f(\vartheta)\, r_1(\vartheta)\,\mathrm{d}\vartheta\right|^2 + \left|\int_{\vartheta_0-2\pi}^{\vartheta_0} \delta f(\vartheta)\, r_1(\vartheta)\,\mathrm{d}\vartheta\right|^2 + \ldots\right\} \tag{II.1.23}$$

und anschließend eine Mittelung über die Gesamtheit von Beschleunigern vorgenommen wurde. Man hat nämlich

$$\langle |F|^2\rangle = \langle \overline{|F|^2}\rangle \tag{II.1.24}$$

($\langle\ \rangle$ = Mittelbildung über die Gesamtheit, $\overline{\ }$ = Mittelbildung über den Umfang), da — bei gleicher Fehlerverteilung der einzelnen Periodizitätsbereiche — in der Gesamtheit auch Beschleuniger vorkommen, die sich von einander nur durch eine zyklische Vertauschung der einzelnen mit Fehlern behafteten Periodizitätsbereiche unterscheiden.

Durch die Tatsache, daß eigentlich zunächst ein Mittelwert über den Umfang (Gl. (II.1.23)) berechnet wurde, wird der für $\hat{r}_{\max}$ maßgebliche Maximalwert von $\left|\int_{\vartheta-2\pi}^{\vartheta} \delta f(\vartheta') r_1(\vartheta')\,\mathrm{d}\vartheta'\right|$ unterschätzt. In diesem Unterabschnitt soll ver-

sucht werden, eine Abschätzung für diese Schwankung und ihren Einfluß auf $|\hat{r}|_{\max}$ zu gewinnen. Wir beschränken uns dabei auf Beschleuniger, bei denen die Fehler in den einzelnen Periodizitätsbereichen unabhängig von einander sind.

Die Größe, die in Gl. (I.1.18) als Faktor von $r_2(\vartheta)$ auftritt und abkürzend mit $\hat{a}_2(\vartheta)$ bezeichnet werden soll, kann in zwei Bestandteile aufgespalten werden

$$\text{(II.1.25)} \qquad \hat{a}_2(\vartheta) = \frac{\int_{\vartheta-2\pi}^{\vartheta} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta'}{W(1-\exp[-2\pi i Q])} =$$

$$= \frac{\int_{0}^{2\pi} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta'}{W(1-\exp[+2\pi i Q])} + \frac{1}{W}\int_{0}^{\vartheta} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta' .$$

Der erste Summand hängt nicht von der Koordinate ϑ ab; der zweite enthält die Schwankung entlang des Umfangs. Falls Q nahe bei einer ganzen Zahl ist, überwiegt der erste Term bei weitem und die Schwankungen werden bedeutungslos. Das ist in Übereinstimmung mit Überlegungen am Schluß von Unterabschn. *a*. Für $\vartheta = 0$ verschwindet der zweite Summand, für $\vartheta = 2\pi$ aber kombiniert er mit dem ersten Summanden zu

$$\text{(II.1.26)} \qquad \hat{a}_2(2\pi) = \frac{\int_{0}^{2\pi} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta'}{W(1-\exp[-2\pi i Q])} = \exp[2\pi i Q]\, a_2(0) ,$$

also zu einer komplexen Zahl vom gleichen Betrag wie $\hat{a}_2(0)$ aber geänderter Phase.

Eine vollständige Untersuchung des Schwankungsgliedes wäre schwierig; deshalb werde eine vereinfachte Abschätzung durchgeführt. Die Abweichung der Größe $|\hat{a}_2(\vartheta)|$ von dem Wert $|\hat{a}_2(0)|$ wird statistisch am größten sein für $\vartheta = \pi$. Der quadratische Mittelwert des zweiten Gliedes ist dort gegeben durch

$$\text{(II.1.27)} \qquad \frac{1}{|W|^2}\left\langle \left| \int_{0}^{\pi} \delta f(\vartheta')\, r_1(\vartheta')\, \mathrm{d}\vartheta' \right|^2 \right\rangle = \frac{M}{2|W|^2} \langle |\varphi|^2 \rangle ;$$

vgl. Gl. (II.1.9). Eigentlich wäre eine genauere Untersuchung der Korrelation der Phasen des ersten und zweiten Terms in Gl. (II.1.25) erforderlich. Der Einfachheit halber soll aber die ungünstige Annahme gemacht werden, daß sich beide komplexen Zahlen bei $\vartheta = \pi$ gerade in Phase addieren. Das würde bedeuten, daß der Faktor $1/|\sin \pi Q|$ in den Gleichungen (II.1.4), (II.1.10), (II.1.15), (II.1.17), (II.1.20) für $\hat{r}_{\max}$ zu ersetzen ist durch $1/\sin \pi Q$ [illegible] 2.

Hierbei ist der zweite Summand aber sicher noch überschätzt, da $|\hat{a}_2(0)|^2$ im allgemeinen kleiner sein wird als der Mittelwert, falls sich beide Summanden in Gl. (II.1.25) für $\vartheta = \pi$ gerade in Phase addieren. Das Resultat ist somit, daß man einen besseren Erwartungswert für die Maximalamplitude der geänderten Gleichgewichtsbahn erhält, falls man Gl. (II.1.10) ersetzt durch [31]

$$\langle \hat{r}_{\max} \rangle = \frac{|r_1|_{\max} \langle |F| \rangle_{W m Q}}{|W|} \left(\frac{1}{|\sin \pi Q|} + \varepsilon \right), \tag{II.1.28}$$

wobei ε eine Zahl der Größenordnung eins ist. Man beachte, daß $1/|\sin \pi Q|$ für halbzahliges Q den Wert eins hat; dort wird man aber nicht arbeiten wegen des Auftretens der instabilen Bereiche (Teil I, Abschn. 2). Für einen Wert von Q, der sich um 1/4 von einer ganzen Zahl unterscheidet, ist $1/|\sin \pi Q| = \sqrt{2}$; das Korrekturglied ε kann dann durchaus von Bedeutung sein.

Es sei nochmals daran erinnert, daß Gl. (II.1.28) keinen wohldefinierten Erwartungswert sondern nur eine Abschätzung darstellt. Trotzdem sollte diese Gleichung die zu erwartende Amplitude der geänderten Gleichgewichtsbahn verhältnismäßig gut wiedergeben. Unsere vereinfachte Abschätzung hat uns nicht erlaubt, eine Aussage über die Abhängigkeit der Größe ε von M und Q zu gewinnen, lediglich ihre Größenordnung (eins) konnte plausibel gemacht werden.

d) Instabile Bereiche. – Bei der statistischen Berechnung der durch Fehler im Feldindex hervorgerufenen instabilen Bereiche (Teil I, Abschn. 2) treten weniger mathematische Komplikationen auf als bei der Untersuchung der geänderten Gleichgewichtsbahn. Dort lagen, wie in den vorangehenden Abschnitten besprochen wurde, die Schwierigkeiten darin, daß sich die Maximalamplitude der geänderten Gleichgewichtsbahn nicht in einfacher geschlossener Form durch die Funktion $\delta f(\vartheta)$ ausdrücken läßt. Bei der Untersuchung der instabilen Bereiche sind, wenigstens bei Beschränkung auf die niedrigste störungstheoretische Näherung, gemäß Gln. (I.2.30), (I.2.31), (I.2.25) einfache Integrale über die Störung $\delta \hat{n}(\vartheta)$ des Feldindex auszuführen.

Zunächst sei bemerkt, daß eine Berechnung von $\delta Q^{(1)}_{\text{mitt}}$, d.h. des Abstandes des Mittelpunktes eines instabilen Bereichs von einem Arbeitspunkt des zugeordneten idealen Beschleunigers mit ganz- oder halbzahligem Q ohne praktisches Interesse ist. Die Verschiebung benachbarter instabiler Bereiche ist nämlich, wie gleich gezeigt werden soll, für großes M nahezu dieselbe; von Bedeutung ist aber allein die Breite des zwischen zwei instabilen Bereichen zur Verfügung stehenden stabilen Gebiets. Die Lage des Mittelpunktes eines

[31] Das Zeichen $\langle\ \rangle_{W m Q}$ ist in Gl. (II.1.11) erklärt.

instabilen Gebiets ist nach Gl. (I.2.31) gegeben durch

$$\delta Q^{(1)}_{\text{mitt}} = \frac{\hat{N}_{12}}{2\pi i W} = \frac{1}{2\pi i W}\int_0^{2\pi} \delta\hat{n}(\vartheta)\,|r_1(\vartheta)|^2\,\mathrm{d}\vartheta\,. \tag{II.1.29}$$

Bei einer Änderung des Arbeitspunktes ändern sich Betrag und Phase von $r_1(\vartheta)$. Bei Übergang von einem Resonanzpunkt zum nächsten sind beide Änderungen für großes M klein; die Phasenänderungen können sich hier (im Gegensatz zu dem Ausdruck für $\delta Q^{(1)}_{\text{inst}}$, siehe Gl. (II.1.31)) nicht auswirken. Damit ist gezeigt, daß die Verschiebungen benachbarter instabiler Bereiche in der Tat für grosses M gleich sind.

Die Breite eines instabilen Bereichs ist nach Gl. (I.2.30) gegeben durch

$$\delta Q^{(1)}_{\text{inst}} = \left|\frac{\hat{N}_{11}}{\pi W}\right|. \tag{II.1.30}$$

Ähnlich wie bei der Behandlung der geänderten Gleichgewichtsbahn werde gesetzt

$$\hat{N}_{11} = \sum_{m=1}^{M} \sigma_m \exp\left[2i(m-1)\mu\right] \tag{II.1.31}$$

mit

$$\sigma_m = \int_{(m-1)\vartheta_0}^{m\vartheta_0} \delta\hat{n}(\vartheta)(r_1(\vartheta)\exp\left[-i(m-1)\mu\right])^2\,\mathrm{d}\vartheta\,. \tag{II.1.32}$$

Es werde wieder die Gleichwahrscheinlichkeit positiver und negativer Fehler postuliert

$$\langle\sigma_m\rangle = \langle\sigma_m^*\rangle = 0\,, \tag{II.1.33}$$

ferner Gleichheit der Fehlerverteilung in den einzelnen Periodizitätsbereichen und statistische Unabhängigkeit der Fehler in verschiedenen Periodizitätsbereichen, also

$$\langle\sigma_m\sigma_l^*\rangle = \delta_{ml}\langle|\sigma|^2\rangle\,. \tag{II.1.34}$$

Man findet damit für den Mittelwert von $|\hat{N}_{11}|^2$

$$\langle|\hat{N}_{11}|^2\rangle = M\langle|\sigma|^2\rangle \tag{II.1.35}$$

und für die Breite des instabilen Bereichs ([31])

$$\langle \delta Q^{(1)}_{\text{inst}} \rangle_{W m Q} = \frac{\sqrt{M}}{\pi |W'|} \; |\sigma| \rangle_{W m Q} \,. \tag{II.1.36}$$

Wiederum tritt die für die statistische Unabhängigkeit der einzelnen Periodizitätsbereiche charakteristische $\sqrt{M}$ auf.

Die weitere Auswertung des Ausdrucks (II.1.36) hängt von der Ursache des Fehlers im Feldindex ab. Nach Unterabschn. *b* der Einleitung kommen insbesondere individuelle Unterschiede der einzelnen Magnete und Aufstellungsfehler entlang der Gleichgewichtsbahn in Frage. Zunächst sollen Magnetfehler betrachtet werden. Setzt man wieder die Aufteilung jedes Periodizitätsbereichs in zwei Sektoren voraus (vgl. die Bemerkungen in Zusammenhang mit Gln. (II.1.12), (II.1.13)), und nimmt man ferner an, daß es physikalisch gerechtfertigt ist, jedem Sektor einen konstanten Fehler $\delta \hat{n}$ zuzuordnen, während die Fehler verschiedener Sektoren statistisch unabhängig sind, so hat man

$$\langle |\sigma| \rangle_{W m Q} = \left(\left| \int\limits_{\text{1. Sektor}} (r_1(\vartheta))^2 \, \mathrm{d}\vartheta \right|^2 + \left| \int\limits_{\text{2. Sektor}} \right|^2 \right)^{\frac{1}{2}} \langle \delta \hat{n} \rangle_{W m Q} \,. \tag{II.1.37}$$

Anwendung der Ähnlichkeitsgesetze von Teil I, Abschn. 3*b* ergibt, daß für ähnliche Anordnungen folgende Gleichung gilt

$$\langle \delta Q^{(1)}_{\text{inst}} \rangle_{W m Q} = n^{\frac{1}{2}} \left\langle \frac{\delta \hat{n}}{n} \right\rangle_{W m Q} f_{\delta n} \left(\frac{n^{\frac{1}{2}}}{M} \right) \tag{II.1.38}$$

mit

$$f_{\delta n} \left(\frac{n^{\frac{1}{2}}}{M} \right) = \left(\frac{n^{\frac{1}{2}}}{M} \right)^{-\frac{1}{2}} \frac{n}{\pi |W'|} \left(\left| \int\limits_{\text{1. Sektor}} (r_1(\vartheta))^2 \mathrm{d}\vartheta \right|^2 + \left| \int\limits_{\text{2. Sektor}} \right|^2 \right)^{\frac{1}{2}} . \tag{II.1.39}$$

Die Größe n stellt wiederum den Absolutwert des Feldindex in einem bestimmten Punkt des Periodizitätsbereichs dar, z.B. den Maximalwert. Die Funktion $f_{\delta n}(n^{\frac{1}{2}}/M)$ wurde für eine Anordnung mit Sektoren gleicher Länge und einem Feldindex gleichen Betrages aber entgegengesetzen Vorzeichens berechnet (Abb. 8). Ihr Unendlichwerden an der oberen Grenze des Stabilitätsbereichs hat denselben Grund wie bei den entsprechenden Funktionen in Unterabschn. *b*.

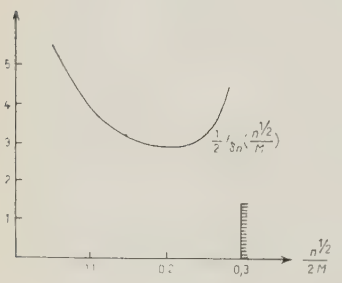


Abb. 8. – Die Funktion $f_{\delta n}(n^{\frac{1}{2}}/M)$, die gemäß Gl. (II.1.38) den Einfluß von eigentlichen Fehlern im Feldindex auf die Breite der instabilen Bereiche angibt.

Aufstellungsfehler entlang der Gleichgewich-

tsbahn können auf einem Unterschied der Länge der einzelnen Sektoren oder auf einem eigentlichen Fehler der Aufstellung beruhen. Es soll zunächst die zweite Möglichkeit betrachtet werden, deren Behandlung nur sinnvoll ist, falls die einzelnen Sektoren durch feldfreie Stücke getrennt sind. Verrückt man einen Sektor um den Winkel $\delta\vartheta$ entlang der Gleichgewichtsbahn, so gilt für die Änderung des Feldindex an einem festen Punkt der Gleichgewichtsbahn

$$\delta n(\vartheta) = -\frac{\mathrm{d}n(\vartheta)}{\mathrm{d}\vartheta}\,\delta\vartheta\,. \tag{II.1.40}$$

Partielle Integration über den Sektor (und etwa das damit verbundene Streufeld im feldfreien Teil) liefert für die auftretenden Integrale zunächst [30]

$$\int \frac{\mathrm{d}n(\vartheta)}{\mathrm{d}\vartheta}\,(r_1(\vartheta))^2\,\mathrm{d}\vartheta = [n(\vartheta)(r_1(\vartheta))^2] - 2\int n(\vartheta)\,r_1(\vartheta)\,r_1'(\vartheta)\,\mathrm{d}\vartheta\,, \tag{II.1.41}$$

wobei der erste Term fortfällt, da die Grenzen des Integrationsbereichs (Sektors) in einem feldfreien Teil ($n=0$) liegen. Anwendung der Bewegungsgleichung und erneute partielle Integration ergeben dann

$$\int \frac{\mathrm{d}n(\vartheta)}{\mathrm{d}\vartheta}\,(r_1(\vartheta))^2\,\mathrm{d}\vartheta = -[(r_1'(\vartheta))^2] \tag{II.1.42}$$

(vgl. Gl. (II.1.19)). Nach den Ähnlichkeitsgesetzen ist dieser Ausdruck, wenn durch W dividiert wird, für ähnliche Anordnungen proportional zu $n^{\frac{1}{2}}$. Für den durch solche Aufstellungsfehler hervorgerufenen Beitrag zu $\delta Q^{(1)}_{\mathrm{inst}}$ erhält man somit

$$\langle \delta Q^{(1)}_{\mathrm{inst}} \rangle_{W,n,Q} = n^{\frac{3}{2}} \langle \delta\vartheta \rangle_{W,n,Q}\, f_{\delta\vartheta}\left(\frac{n^{\frac{1}{2}}}{M}\right), \tag{II.1.43}$$

wobei die Funktion [30]

$$f_{\delta\vartheta}\left(\frac{n^{\frac{1}{2}}}{M}\right) = \left(\frac{n^{\frac{1}{2}}}{M}\right)^{-\frac{1}{2}} \frac{1}{\pi\,|W|\,n^{\frac{1}{2}}} \left(\left| \int_{1.\,\mathrm{Sektor}} [(r_1'(\vartheta))^2]\,\mathrm{d}\vartheta \right|^2 + \left| \int_{2.\,\mathrm{Sektor}} \right|^2 \right)^{\frac{1}{2}} \tag{II.1.44}$$

wiederum nur von $n^{\frac{1}{2}}/M$ abhängt. Übrigens hat man nach Gl. (II.1.18)

$$n^{\frac{1}{2}}\,\delta\vartheta = 2\pi\,\frac{n^{\frac{1}{2}}}{M}\,\frac{\delta l}{L}\,, \tag{II.1.45}$$

wobei δl die geometrische Verrückung eines Sektors und L wiederum die Periodenlänge darstellt.

Auch diese Funktion wurde für die Anordnung mit zwei Sektoren gleicher Länge und mit Feldindizes gleichen Betrages aber entgegengesetzten Vorzeichens berechnet. Das ist eigentlich nicht ganz sinnvoll, da bei dieser Anordnung angenommen ist, daß die Sektoren unmittelbar an einander stossen und unabhängige Verrückungen daher nicht möglich sind. Wenn trotzdem Gl. (II.1.44) benutzt wurde, so hat man das Ergebnis so zu interpretieren, daß man zwischen den einzelnen Sektoren kleine feldfreie Bereiche annimmt, die

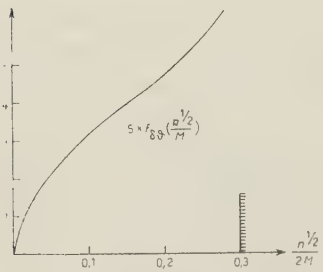


Abb. 9. – Die Funktion $f_{\delta\vartheta}(n^{\frac{1}{2}}/M)$, die gemäß Gl. (II.1.43) den Einfluß tangentialer Aufstellungsfehler der Sektoren auf die instabilen Bereiche angibt.

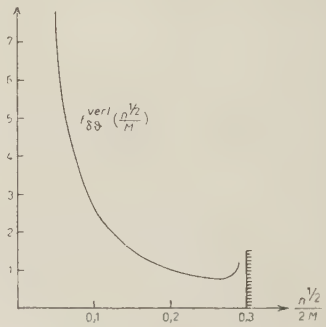


Abb. 10. – Die Funktion $f_{\delta\vartheta}^{\mathrm{verl}}(n^{\frac{1}{2}}/M)$, die gemäß Gl. (II.1.43) den Einfluß einer Verlängerung (oder Verkürzung) der Sektoren auf die instabilen Bereiche angibt.

den Verlauf der Betatronschwingungen noch nicht wesentlich ändern sollen. Die Größenordnung des Ergebnisses ist zugleich kennzeichnend für die Ergebnisse, die man bei anderen Anordnungen zu erwarten hat. Die Funktion $f_{\delta\vartheta}(n^{\frac{1}{2}}/M)$ ist in Abb. 9 dargestellt.

Von stärkerem Einfluß als tangentiale Aufstellungsfehler sind Fehler in der Länge der einzelnen Sektoren. Es gelten wieder Formeln ähnlich Gl. (II.1.43), wobei $\delta\vartheta$ jetzt die Verlängerung (bzw. Verkürzung) eines Sektors bedeutet. Für die Anordnung, die allen zahlenmäßigen Rechnungen dieser Arbeit zu Grunde liegt, wurde die Funktion $f_{\delta\vartheta}^{\mathrm{verl}}(n^{\frac{1}{2}}/M)$ berechnet und in Abb. 10 dargestellt.

2. – Wahrscheinlichkeitsverteilungen.

a) *Geänderte Gleichgewichtsbahn.* – In diesem Abschnitt werden die statistischen Untersuchungen auf die Wahrscheinlichkeitsverteilung einiger Größen ausgedehnt. Die Ermittlung einer Wahrscheinlichkeitsverteilung soll zunächst für die Amplitude der geänderten Gleichgewichtsbahn durchgeführt werden. Wegen der schon besprochenen Schwierigkeiten (vgl. Schluß von Abschn. 1*a*) soll sich die Untersuchung auf die Größe F (Gl. II.1.3) beschränken. Wiederum

soll die Annahme der Gleichwahrscheinlichkeit positiver und negativer Fehler und der Unabhängigkeit der Fehler in verschiedenen Periodizitätsbereichen gemacht werden.

In Gl. (II.1.5) sind die Größen q_m dann unabhängig von einander. Trägt man die auf der rechten Seite der Gleichung stehende Summe in der komplexen Ebene auf, so erhält man einen Streckenzug, wobei die Längen der einzelnen Strecken unabhängig von einander sind und alle dem gleichen statistischen Gesetz gehorchen. Man hat also ein typisches Irrfahrtproblem in zwei Dimensionen. Für die Länge des Vektors F (im Sinne des in der komplexen Ebene üblichen Sprachgebrauchs) wird man also eine zweidimensionale Gaußverteilung (Raleighverteilung) der Form

$$w(|F|)\,\mathrm{d}|F| = \frac{2}{\langle |F|^2 \rangle} \exp\left[-\frac{|F|^2}{\langle |F|^2 \rangle}\right] |F|\,\mathrm{d}|F| \tag{II.2.1}$$

erwarten. Die Größe $\langle |F|^2 \rangle$ bedeutet den quadratischen Mittelwert von F. Die Wahrscheinlichkeitsverteilung (II.2.1) ist zusammen mit der integrierten Wahrscheinlichkeitsverteilung

$$W(|F|) = \int_0^{|F|} w(|F'|)\,\mathrm{d}|F'| = 1 - \exp\left[-\frac{|F|^2}{\langle |F|^2 \rangle}\right] \tag{II.2.2}$$

in Abb. 11 dargestellt. Der Beweis für die (näherungsweise) Gültigkeit von Gl. (II.2.1) wird gleich gegeben werden. Zunächst aber sei bemerkt, daß nach dieser Gleichung die Größe der zu erwartenden Werte von $|F|$ recht gut durch den in Abschnitt 1*b* berechneten Wert von $\langle |F| \rangle_{WmQ}$ gekennzeichnet ist. In etwa 60% aller Fälle ist $|F| \leqslant \langle |F| \rangle_{WmQ}$, in nur 10% aller Fälle hat man $|F| \geqslant 1{,}5 \cdot \langle |F| \rangle_{WmQ}$, und es ist sehr unwahrscheinlich (weniger als 2% aller Fälle), einen Wert von $|F|$ zu finden, der größer ist als $2\langle |F| \rangle_{WmQ}$.

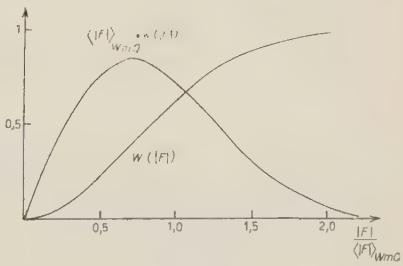


Abb. 11. – Differentielle und integrierte Wahrscheinlichkeitsverteilung der Größe $|F|$ nach Gln. (II.2.1), (II.2.2). Die Größe $\delta Q_{\mathrm{inst}}^{(1)}$ gehorcht nach Gl. (II.2.7) den gleichen Wahrscheinlichkeitsverteilungen.

Der Beweis für die Gültigkeit der Wahrscheinlichkeitsverteilung soll durch Berechnung der Momente von $|F|$, d.h. der Mittelwerte $\langle |F|^{2k} \rangle$ der Potenzen erfolgen, wobei man sich auf die geraden Potenzen beschränken darf. Es wird sich zeigen, daß die Momente $\langle |F|^{2k} \rangle$ für nicht zu großes k ($k \ll M$) näherungsweise mit den Momenten übereinstimmen, die sich unter Benutzung von (II.2.1) ergeben. Die Tatsache, daß die unmittelbar und die nach Gl. (II.2.1)

berechneten Momente nur für nicht zu großes k, und auch dann nur näherungsweise, übereinstimmen, bedeutet, daß die Wahrscheinlichkeitsverteilung (II.2.1) auch nur näherungsweise für nicht zu großes F erfüllt ist. Bei genügend großem Wert von M machen sich die Unterschiede aber erst in einem Gebiet bemerkbar, in dem sowohl die wahre als auch die durch Gl. (II.2.1) gegebene Wahrscheinlichkeitsverteilung bereits vernachlässigbar klein sind.

Das Moment $\langle|F|^2\rangle$ wurde bereits in Gl. (II.1.9) berechnet. Dabei wurde wesentlich Gebrauch gemacht von Gl. (II.1.7). (Gleichwahrscheinlichkeit positiver und negativer Fehler) und Gl. (II.1.8) (gleiche Fehlerverteilung in allen Periodizitätsbereichen, jedoch statistische Unabhängigkeit der Fehler in verschiedenen Periodizitätsbereichen). In Verallgemeinerung von Gl. (II.1.7) soll gefordert werden, daß alle Mittelwerte mit einer ungeraden Anzahl von φ-Faktoren verschwinden, während in Analogie zu Gl. (II.1.8) beispielsweise gilt

$$\text{(II.2.3)}\qquad \langle\varphi_l\varphi_{l'}\varphi_m^*\varphi_{m'}^*\rangle = (\delta_{lm}\delta_{l'm'} + \delta_{lm'}\delta_{l'm} - 2\delta_{lm}\delta_{ll'}\delta_{mm'})\langle|\varphi|^2\rangle^2 + \\ + (\delta_{ll'}\delta_{mm'} - \delta_{lm}\delta_{ll'}\delta_{mm'})\,|\langle\varphi^2\rangle|^2 + \delta_{lm}\delta_{ll'}\delta_{mm'}\langle|\varphi|^4\rangle\,.$$

Damit erhält man für das Moment

$$\text{(II.2.4)}\qquad \langle|F|^4\rangle = \sum_{l,l',m,m'=1}^{M} \exp[i(l+l'-m-m')\mu]\langle\varphi_l\varphi_{l'}\varphi_m^*\varphi_{m'}^*\rangle = \\ = 2M(M-1)\langle|\varphi|^2\rangle^2 + \left(\frac{1-\cos 4\pi Q}{1-\cos 2\mu} - M\right)|\langle\varphi^2\rangle|^2 + M\langle|\varphi|^4\rangle\,.$$

Hieraus kann man ablesen, daß näherungsweise gilt

$$\text{(II.2.5)}\qquad \langle|F|^4\rangle \approx 2M^2\,\langle|\varphi|^2\rangle^2 = 2\,\langle|F|^2\rangle^2$$

(vgl. Gl. (II.1.9)), falls $M \gg 1$ und falls alle Mittelwerte $\langle|\varphi|^2\rangle^2$, $|\langle\varphi^2\rangle|^2$, $\langle|\varphi|^4\rangle$ etwa die gleiche Größenordnung besitzen. Die Größe $(1-\cos 4\pi Q)/(1-\cos 2\mu)$ ist stets endlich und geht an den Grenzen eines Stabilitätsbereichs gegen null. In Verallgemeinerung der zu Gl. (II.2.5) führenden Rechnung findet man, daß gilt

$$\text{(II.2.6)}\qquad \langle|F|^{2k}\rangle \approx k!\,\langle|F|^2\rangle^k\,,$$

solange $k \ll M$ ist. Diese Beziehung ist aber gerade kennzeichnend für die Wahrscheinlichkeitsverteilung (II.2.1).

b) Instabile Bereiche. – Für die Größe $\delta Q^{(1)}_{\text{inst}}$ (vgl. Gl. (I.2.30)), die für die Breite eines instabilen Bereichs kennzeichnend ist, verlaufen die Rechnungen genau so wie für F im vorangehenden Unterabschnitt. Statt Gl. (II.1.5) ist jetzt Gl. (II.1.31) zu Grunde zu legen, die wiederum als zweidimensionale

Irrfahrt gedeutet werden kann. Man erwartet also die Wahrscheinlichkeitsverteilung

$$w(\delta Q^{(1)}_{\text{inst}})\,\mathrm{d}(\delta Q^{(1)}_{\text{inst}}) = \frac{2}{\delta Q^{(1)2}_{\text{inst}}\rangle} \exp\left[-\frac{\delta Q^{(1)2}_{\text{inst}}}{\langle \delta Q^{(1)2}_{\text{inst}}\rangle} \right] \delta Q^{(1)}_{\text{inst}}\,\mathrm{d}(\delta Q^{(1)}_{\text{inst}}) \,; \tag{II.2.7}$$

vgl. Abb. 11. Sie ergibt sich aus einer Berechnung der Momente von $\delta Q^{(1)}_{\text{inst}}$. Entsprechend Gl. (II.2.3) hat man

$$\begin{aligned}\langle \sigma_l \sigma_{l'} \sigma^*_m \sigma^*_{m'} \rangle = {} & (\delta_{lm}\delta_{l'm'} + \delta_{lm'}\delta_{l'm} - 2\delta_{lm}\delta_{ll'}\delta_{mm'}) \langle |\sigma|^2\rangle^2 + \\ & + (\delta_{ll'}\delta_{mm'} - \delta_{lm}\delta_{ll'}\delta_{mm'}) |\langle \sigma^2\rangle|^2 + \delta_{lm}\delta_{ll'}\delta_{mm'} \langle |\sigma|^4\rangle\end{aligned} \tag{II.2.8}$$

und damit

$$\langle |\hat{N}_{11}|^4\rangle = 2M(M-1)\langle |\sigma|^2\rangle^2 + \left(\frac{1 - \cos 8\pi Q}{1 - \cos 4\mu} - M \right) |\langle \sigma^2\rangle|^2 + M \langle |\sigma|^4\rangle \,, \tag{II.2.9}$$

wobei aber $(1 - \cos 8\pi Q)/(1 - \cos 4\mu)$ wegen des streng ganz- bzw. halbzahligen Wertes von Q exakt verschwindet. Allgemein findet man für $k \ll M$

$$\langle |\hat{N}_{11}|^{2k}\rangle \approx k!\, M^k \langle |\sigma|^2\rangle^{2k} = k! \langle |\hat{N}_{11}|^2\rangle \,, \tag{II.2.10}$$

was für die Wahrscheinlichkeitsverteilung (II.2.7) kennzeichnend ist.

Von größerem praktischem Interesse als die Breite eines einzelnen instabilen Bereichs ist diejenige des stabilen Bereichs zwischen zwei benachbarten instabilen Bereichen, die wegen der (näherungsweise) gleichen Verschiebung benachbarter instabiler Bereiche (siehe Anfang von Abschn. 1*d*) gegeben ist durch

$$\delta Q_{\text{stab}} = \tfrac{1}{2}(1 - (\delta Q^{(1)}_{\text{inst}})_Q - (\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}}) \,. \tag{II.2.11}$$

Hierbei bedeuten $(\delta Q^{(1)}_{\text{inst}})_Q$ und $(\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}}$ die Breite der instabilen Bereiche, wenn die Zahl der Betatronwellenlängen auf dem Umfang gleich Q und wenn sie gleich $Q + \frac{1}{2}$ ist. Es soll deshalb die Wahrscheinlichkeitsverteilung der Größe

$$\Delta = (\delta Q^{(1)}_{\text{inst}})_Q + (\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}} \tag{II.2.12}$$

berechnet werden, die allgemein gegeben ist durch [32]

$$w(\Delta) = \iint w((\delta Q^{(1)}_{\text{inst}})_Q, (\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}})\, \delta(\Delta - (\delta Q^{(1)}_{\text{inst}})_Q - (\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}})\, \mathrm{d}(\delta Q^{(1)}_{\text{inst}})_Q\, \mathrm{d}(\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}} \,. \tag{II.2.13}$$

[32] $\delta(x)$ ist die Diracsche Deltafunktion.

Weiter unten soll gezeigt werden, daß die Breiten benachbarter instabiler Bereiche für großes M von einander praktisch statistisch unabhängig sind, daß also gilt

$$w\left((\delta Q^{(1)}_{\text{inst}})_Q, (\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}}\right) \approx w\left((\delta Q^{(1)}_{\text{inst}})_Q\right) w\left((\delta Q^{(1)}_{\text{inst}})_{Q+\frac{1}{2}}\right) . \tag{II.2.14}$$

Hier ist für die Wahrscheinlichkeitsverteilungen auf der rechten Seite Gl. (II.2.7) einzusetzen, und zwar mit in beiden Fällen gleichem $\langle \delta Q^{(1)2}_{\text{inst}} \rangle$. Eine einfache Rechnung ergibt für die Wahrscheinlichkeitsverteilung (II.2.13) [33]

$$\langle \delta Q^{(1)}_{\text{inst}} \rangle_{W m Q} w(\Delta) = \frac{\Delta}{\langle \delta Q^{(1)}_{\text{inst}} \rangle_{W m Q}} \exp\left[- \frac{\Delta^2}{\langle \delta Q^{(1)2}_{\text{inst}} \rangle} \right] + \sqrt{\frac{\pi}{2}} \left| \frac{\Delta^2}{\langle \delta Q^{(1)2}_{\text{inst}} \rangle} - 1 \right| \exp\left[- \frac{\Delta^2}{2 \langle \delta Q^{(1)2}_{\text{inst}} \rangle} \right] \Phi\left(\frac{\Delta}{\sqrt{2} \langle \delta Q^{(1)}_{\text{inst}} \rangle_{W m Q}} \right) \tag{II.2.15}$$

und für die integrierte Wahrscheinlichkeitsverteilung

$$W(\Delta) = \int_0^{\Delta} w(\Delta') \, \mathrm{d}\Delta' \tag{II.2.16}$$

das folgende Resultat [33]

$$W(\Delta) = 1 - \exp\left[- \frac{\Delta^2}{\langle \delta Q^{(1)2}_{\text{inst}} \rangle} \right] - \sqrt{\frac{\pi}{2}} \frac{\Delta}{\langle \delta Q^{(1)}_{\text{inst}} \rangle_{W m Q}} \exp\left[- \frac{\Delta^2}{2 \langle \delta Q^{(1)2}_{\text{inst}} \rangle} \right] \Phi\left(\frac{\Delta}{\sqrt{2} \langle \delta Q^{(1)}_{\text{inst}} \rangle_{W m Q}} \right) . \tag{II.2.17}$$

Beide Wahrscheinlichkeitsverteilungen sind in Abb. 12 aufgetragen. Die ganze Rechnung ist natürlich nur sinnvoll für Breiten Δ, die klein gegen eins sind.

Es ist noch der Beweis für die statistische Unabhängigkeit der Breiten benachbarter instabiler Bereiche nachzutragen. Es kann etwas allgemeiner gezeigt werden, daß die

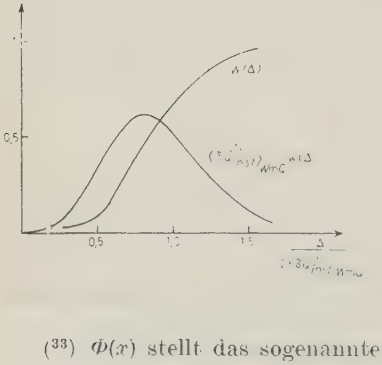


Abb. 12. – Differentielle und integrierte Wahrscheinlichkeitsverteilung für die Summe Δ der Breiten benachbarter instabiler Bereiche nach Gln. (II.2.15), (II.2.17).

[33] $\Phi(x)$ stellt das sogenannte Fehlerintegral dar. $\Phi(x) = (2/\sqrt{\pi}) \int_0^x \mathrm{e}^{-y^2} \mathrm{d}y$; vgl. etwa E. JAHNKE u. F. EMDE: *Funktionentafeln* (Leipzig und Berlin, 1938).

Breiten von instabilen Bereichen, die (bei Q und Q') nicht zu weit von einander entfernt liegen, statistisch unabhängig von einander sind. Es ist hierfür zu beweisen

$$\langle |\hat{N}_{11}^{(Q)}|^{2k} |\hat{N}_{11}^{(Q')}|^{2k'} \rangle \approx \langle |\hat{N}_{11}^{(Q)}|^{2k} \rangle \langle |\hat{N}_{11}^{(Q')}|^{2k'} \rangle \,. \tag{II.2.18}$$

Dabei sei angenommen, daß die einzelnen Beiträge σ_m sich bei Übergang von $N_{11}^{(Q)}$ zu $N_{11}^{(Q')}$ praktisch nicht ändern und daß der Unterschied beider Größen allein von der Änderung der Phase μ herrührt. Gl. (II.2.18) soll explizit vorgerechnet werden für $k = k' = 1$. Unter Verwendung von Gl. (II.2.8) erhält man

$$\begin{aligned} |\hat{N}_{11}^{(Q)}|^2 |\hat{N}_{11}^{(Q')}|^2 \rangle = &\left(M^2 - \frac{1 - \cos 4\pi(Q - Q')}{1 - \cos 2(\mu - \mu')} - 2M \right) \langle |\sigma|^2 \rangle^2 + \\ &+ \left(\frac{1 - \cos 4\pi(Q + Q')}{1 - \cos 2(\mu + \mu')} - M \right) |\langle \sigma^2 \rangle|^2 - M^2 \langle \sigma^4 \rangle \,. \end{aligned} \tag{II.2.19}$$

Hierbei verschwindet $(1 - \cos 4\pi(Q+Q'))/(1 - \cos 2(\mu+\mu'))$ identisch, während $(1 - \cos 4\pi(Q - Q'))/(1 - \cos 2(\mu - \mu'))$ verschwindet, falls Q und Q' wirklich verschieden sind. Für großes M ergibt sich die rechte Seite von Gl. (II.2.19) in der Tat gleich

$$M^2 \langle |\sigma|^2 \rangle = \langle |\hat{N}_{11}^{(Q)}|^2 \rangle \langle |\hat{N}_{11}^{(Q')}|^2 \rangle \,. \tag{II.2.20}$$

Entsprechend schließt man für größere k und k'.

3. – Zusammenfassung von Teil II.

Um die Störungen abschätzen zu können, die in einem Synchrotron mit starker Stabilisierung zu erwarten sind, wurde in diesem Teil eine Auswertung der in Teil I abgeleiteten Ergebnisse unter statistischen Annahmen für die verschiedenen Fehler durchgeführt. Speziell für eine Anordnung, bei der jeder Periodizitätsbereich aus zwei Sektoren gleicher Länge mit Feldindizes gleichen Betrages aber entgegengesetzen Vorzeichens besteht, wurden die Ergebnisse quantitativ in Form von Kurven (Abb. 7 bis 10) angegeben.

Bei Untersuchung der geänderten Gleichgewichtsbahn ließ sich eine etwas gröbere Abschätzung nicht vermeiden (Abschn. 1*a*), auch die Schwankung der für die Amplitude der Gleichgewichtsbahn kennzeichnenden Größe entlang des Umfangs wurde nur abgeschätzt (Abschn. 1*c*). Die eigentlich quantitative Behandlung dieser Größe erfolgte in Abschnitt 1*b*, während ihre Wahrscheinlichkeitsverteilung in der gedachten Gesamtheit in Abschnitt 2*a* untersucht wurde.

Bei Untersuchung der instabilen Bereiche ist die Verschiebung der Mittel-

punkte dieser Bereiche ohne Interesse, da benachbarte instabile Bereiche praktisch gleich stark verschoben werden. Der quadratische Mittelwert der Breite eines instabilen Bereichs wurde in Abschnitt 1*c*, die Wahrscheinlichkeitsverteilung dieser Größe in Abschnitt 2*b* behandelt. In Abschnitt 2*b* wurde ferner die Wahrscheinlichkeitsverteilung für die Summe der Breiten benachbarter instabiler Bereiche untersucht; aus ihr läßt sich leicht die eigentlich interessierende Wahrscheinlichkeitsverteilung für die Breite des stabilen Bereichs zwischen zwei instabilen Bereichen gewinnen.

* * *

Es ist dem Verfasser ein Bedürfnis, den Herren Prof. N. Bohr und Prof. W. Heisenberg für ihr freundliches Interesse und für anregende Diskussionen vielmals zu danken. Die Klärung der behandelten Probleme wurde wesentlich gefördert durch zahlreiche Gespräche mit Mitarbeitern der Theoretischen Studiengruppe und der Protonen-Synchrotron-Gruppe der Europäischen Organisation für Kernphysikalische Forschung (CERN) und mit Angehörigen des Instituts für Theoretische Physik der Universität Kopenhagen, des Max Planck-Instituts für Physik, Göttingen, und des Physikalischen Instituts der Universität Bonn. Ihnen allen sei auch an dieser Stelle sehr gedankt.

Anhang 1.

Bewegungsgleichungen in linearer Näherung.

Die in der Einleitung, Gln. (E.2) und (E.3), gegebenen Bewegungsgleichungen für ein Teilchen im Synchrotron sind nicht neu. Aus den bisher gegebenen Ableitungen ([34]) wird aber nicht ganz klar, ob die Gleichungen ohne Zusatzglieder auch gültig sind, wenn Führungsfeld und Feldindex nicht entlang der Gleichgewichtsbahn konstant sind. Tatsächlich wird sich zeigen, daß derartige Zusatzglieder zunächst durchaus auftreten, sich in den endgültigen Bewegungsgleichungen aber gerade fortheben. Aus diesem Grunde ist die nachfolgende Ableitung der Bewegungsgleichungen vielleicht nicht ganz ohne Interesse.

a) Vernachlässigung der Beschleunigung. – Zunächst werde die Wirkung der Beschleunigungsspannung vernachlässigt, das Problem also fiktiv als ein stationäres behandelt. $\tilde{\boldsymbol{R}}(t)$ stelle den Radiusvektor einer möglichen Teilchenbahn, dar, die der Einfachheit halber in einer festen Ebene liegen möge. Für ein Teilchen richtiger Energie wählt man für diese Bahn zweckmässig die in

([34]) D. W. Kerst und R. Serber: *Phys. Rev.*, **60**, 53 (1941); N. M. Blachman und E. D. Courant: *Rev. Sci. Instr.*, **20**, 596 (1949).

sich zurücklaufende Gleichgewichtsbahn. Auf der Bahn $\tilde{\boldsymbol{R}}(t)$ läuft das Teilchen mit konstanter Geschwindigkeit v um, so daß für die zeitliche Ableitung gilt

$$\dot{\tilde{\boldsymbol{R}}}(t) = v\,\boldsymbol{t}_0(t)\,. \tag{A.1.1}$$

Der Vektor $\boldsymbol{t}_0$, der seinerseits natürlich von der Zeit t abhängt, bedeutet den Tangenteneinheitsvektor an die Bahn. Es werde ein entlang der Bahn $\tilde{\boldsymbol{R}}(t)$ umlaufendes begleitendes Dreibein eingeführt, das aus dem Vektor $\boldsymbol{t}_0(t)$, dem auf der Bahnebene senkrechten Einheitsvektor $\boldsymbol{z}_0$ und der in der Bahnebene liegenden Normalen $\boldsymbol{r}_0(t)$ besteht (Abb. 13).

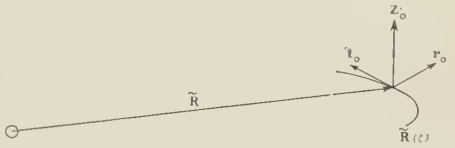


Abb. 13. – Der Vektor $\tilde{\boldsymbol{R}}$ und die Achsen des mitbewegten Dreibeins.

Unter Benutzung rechtwinkliger kartesischer Koordinaten τ, z, r in dem begleitenden Dreibein kann jeder Punkt $\boldsymbol{R}$ dargestellt werden in der Form

$$\boldsymbol{R} = \tilde{\boldsymbol{R}}(t) + \boldsymbol{t}_0(t)\tau + \boldsymbol{z}_0 z + \boldsymbol{r}_0(t) r\,. \tag{A.1.2}$$

Die folgenden Rechnungen sollen sich auf die Umgebung der Bahn $\tilde{\boldsymbol{R}}(t)$ beschränken; höhere als erste Potenzen der Koordinaten τ, z, r sollen daher stets fortgelassen werden.

Für eine umlaufendes Teilchen sind τ, z, r Funktionen der Zeit. Unter Benutzung von

$$\dot{\boldsymbol{t}}_0 = -\boldsymbol{r}_0 \frac{v}{\varrho}\,, \qquad \dot{\boldsymbol{r}}_0 = \boldsymbol{t}_0 \frac{v}{\varrho} \tag{A.1.3}$$

mit ϱ als Krümmungsradius der Bahn $\tilde{\boldsymbol{R}}$ an der betreffenden Stelle findet man für die zeitlichen Ableitungen von $\boldsymbol{R}$

$$\begin{cases} \dot{\boldsymbol{R}} = \boldsymbol{t}_0 \left\{ \dot{\tau} + v\left(1 + \frac{r}{\varrho}\right) \right\} + \boldsymbol{z}_0 \dot{z} + \boldsymbol{r}_0 \left\{ \dot{r} - \frac{v\tau}{\varrho} \right\}, \\ \ddot{\boldsymbol{R}} = \boldsymbol{t}_0 \left\{ \ddot{\tau} - \left(\frac{v}{\varrho}\right)^2 \tau + v\left(2\frac{\dot{r}}{\varrho} + r\left(\frac{1}{\varrho}\right)^{\cdot}\right) \right\} + \boldsymbol{z}_0 \ddot{z} + \\ \quad + \boldsymbol{r}_0 \left\{ \ddot{r} - \frac{v^2}{\varrho}\left(1 + \frac{r}{\varrho}\right) - 2\frac{v}{\varrho}\dot{\tau} - r\left(\frac{1}{\varrho}\right)^{\cdot}\tau \right\}. \end{cases} \tag{A.1.4}$$

Für später (Gl. (A.1.13)) ist die Bemerkung wichtig, daß derselbe feste Raumpunkt wegen der Bewegung des Dreibeins zu verschiedenen Zeiten verschiedene Koordinaten hat. Aus der Bedingung

$$\boldsymbol{R}(t + \mathrm{d}t,\ \tau + \mathrm{d}\tau,\ z + \mathrm{d}z,\ r + \mathrm{d}r) = \boldsymbol{R}(t, \tau, z, r) \tag{A.1.5}$$

folgt unter Benutzung der ersten Gl. (A.1.4)

$$\mathrm{d}\tau = -v\,\mathrm{d}t\left(1 + \frac{r}{\varrho}\right), \qquad \mathrm{d}z = 0\,, \qquad \mathrm{d}r = v\,\mathrm{d}t\frac{\tau}{\varrho}\,, \tag{A.1.6}$$

was sich für $r, \tau \to 0$ zu

$$\mathrm{d}\tau = -v\,\mathrm{d}t\,, \qquad \mathrm{d}z = \mathrm{d}r = 0 \tag{A.1.7}$$

reduziert.

Das Magnetfeld $\boldsymbol{H}(\boldsymbol{R})$ in der Umgebung eines Punktes $\tilde{\boldsymbol{R}}$ werde dargestellt

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{R}) = \boldsymbol{z}_0 H_z + \boldsymbol{t}_0(K_{11}\tau + K_{12}z + K_{13}r) + \\ + \boldsymbol{z}_0(K_{21}\tau + K_{22}z + K_{23}r) + \\ + \boldsymbol{r}_0(K_{31}\tau + K_{32}z + K_{33}r) + \text{quadr. Glieder.}\end{aligned} \tag{A.1.8}$$

Dabei hängen H_z und die Koeffizienten K_{jl} nur von t (wegen der Bewegung des Dreibeins), nicht aber von τ, r, z ab. Es ist beispielsweise

$$K_{22} = \frac{\partial H_z}{\partial z}\,, \quad K_{23} = \frac{\partial H_z}{\partial r}\,. \tag{A.1.9}$$

Aus den Bedingungen

$$\operatorname{div} \boldsymbol{H} = 0\,, \qquad \operatorname{rot} \boldsymbol{H} = 0 \tag{A.1.10}$$

folgt für die Koeffizienten K_{jl}

$$K_{11} + K_{22} + K_{33} = 0\,, \qquad K_{jl} = K_{lj}\,. \tag{A.1.11}$$

Ein zeitlich konstantes Magnetfeld wird zu verschiedenen Zeiten durch verschiedene H_z, K_{jl} usw. dargestellt. Aus

$$\boldsymbol{H}(t + \mathrm{d}t,\ \tau + \mathrm{d}\tau,\ z + \mathrm{d}z,\ r + \mathrm{d}r) = \boldsymbol{H}(t, \tau, z, r) \tag{A.1.12}$$

und Gl. (A.1.7) folgt

$$\boldsymbol{z}_0 \dot{H}_z - v(\boldsymbol{t}_0 K_{11} + \boldsymbol{z}_0 K_{21} + \boldsymbol{r}_0 K_{31}) + \text{kleine Glieder} = 0\,, \tag{A.1.13}$$

wobei die «kleinen Glieder» in τ, z, r mindestens von erster Ordnung sind. Aus dieser Gleichung erhält man (unter Mitbenutzung von Gl. (A.1.11))

$$\begin{cases} K_{21}v = K_{12}v = \dot{H}_z\,, & K_{11} = K_{31} = K_{13} = 0\,, \\ K_{22} + K_{33} = 0\,, & K_{23} = K_{32}\,. \end{cases} \tag{A.1.14}$$

Jetzt kann in der rechten Seite der Bewegungsgleichung

$$(m\dot{\boldsymbol{R}})^{\cdot} = \frac{e}{c}\,\dot{\boldsymbol{R}} \times \boldsymbol{H} \tag{A.1.15}$$

das Vektorprodukt $\dot{\boldsymbol{R}} \times \boldsymbol{H}$ errechnet werden. Bei Beschränkung auf lineare

Glieder findet man

$$\text{(A.1.16)}\qquad \dot{\boldsymbol{R}}\times\boldsymbol{H}=\boldsymbol{t}_0\left\{\frac{v\tau}{\varrho}-\dot{r}\right\}H_z+\boldsymbol{z}_0\{-vK_{32}z-vK_{33}r\}+$$
$$+\boldsymbol{r}_0\left\{H_z v\left(1+\frac{r}{\varrho}\right)+\dot{H}_z\tau+H_z\dot{\tau}+vK_{22}z+vK_{23}r\right\}$$

und somit unter Benutzung von Gl. (A.1.4) und Beachtung von

$$\text{(A.1.17)}\qquad \dot{m}=0$$

die drei skalaren Bewegungsgleichungen

$$\text{(A.1.18)}\qquad \begin{cases} \ddot{\tau}-\left(\frac{v}{\varrho}\right)^2\tau+v\left(2\frac{\dot{r}}{\varrho}+r\left(\frac{1}{\varrho}\right)^{\cdot}\right)=\frac{eH_z}{mc}\left(\frac{v\tau}{\varrho}-\dot{r}\right), \\ \ddot{z}=-\frac{ev}{mc}(K_{32}z+K_{33}r), \\ \ddot{r}-\frac{v^2}{\varrho}\left(1+\frac{r}{\varrho}\right)-2\frac{v}{\varrho}\dot{\tau}-v\tau\left(\frac{1}{\varrho}\right)^{\cdot}=\frac{e}{mc}\left(H_z v\left(1+\frac{r}{\varrho}\right)+\right. \\ \qquad\qquad\qquad\qquad\left.+\dot{H}_z\tau+H_z\dot{\tau}+v(K_{22}z+K_{33}r)\right). \end{cases}$$

Hierbei bedeutet m die der konstanten Geschwindigkeit v entsprechende relativistische Masse.

Nunmehr wird von der Voraussetzung Gebrauch gemacht, daß $\tilde{\boldsymbol{R}}(t)$ eine Bahn darstellt, die unter dem Einfluß des Magnetfeldes $\boldsymbol{H}(\boldsymbol{R})$ wirklich durchlaufen werden kann, d.h. $\tau=z=r=0$ soll eine Lösung des Systems (A.1.18) sein. Aus der ersten oder dritten Gleichung findet man die Bedingung (E.1)

$$\text{(A.1.19)}\qquad H_z=-\frac{pc}{e\varrho},\quad \dot{H}_z=-\frac{pc}{e}\left(\frac{1}{\varrho}\right)^{\cdot}.$$

Damit vereinfachen sich die Bewegungsgleichungen in folgender Weise

$$\text{(A.1.20)}\qquad \begin{cases} \ddot{\tau}+v\left(\frac{r}{\varrho}\right)^{\cdot}=0, \\ \ddot{z}=\frac{v^2}{H_z\varrho}\left(z\frac{\partial H_z}{\partial r}-r\frac{\partial H_z}{\partial z}\right), \\ \ddot{r}-\frac{v\dot{\tau}}{\varrho}=-\frac{v^2}{H_z\varrho}\left(z\frac{\partial H_z}{\partial z}+r\frac{\partial H_z}{\partial r}\right). \end{cases}$$

Die Geschwindigkeit v eines Teilchens im Magnetfeld ist zeitlich konstant (für das auf der Gleichgewichtsbahn umlaufende Teilchen wurde hiervon bereits in Gl. (A.1.1) Gebrauch gemacht); in linearer Näherung gilt als Folge von Gl. (A.1.4) also

$$\text{(A.1.21)}\qquad \dot{\tau}+v\frac{r}{\varrho}=0.$$

Diese Gleichung stellt ersichtlich die erste Gl. (A.1.20) in integrierter Form dar.

Diese erste Gleichung kann also fortgelassen werden, während sich die beiden anderen unter Verwendung von Gl. (A.1.21) folgendermaßen schreiben

$$\text{(A.1.22)} \qquad \begin{aligned} \ddot{z} &= \frac{v^2}{H_z \varrho}\left(z\,\frac{\partial H_z}{\partial r} - r\,\frac{\partial H_z}{\partial z}\right), \\ \ddot{r} + \left(\frac{v}{\varrho}\right)^2 r &= -\frac{v^2}{H_z \varrho}\left(z\,\frac{\partial H_z}{\partial z} + r\,\frac{\partial H_z}{\partial r}\right). \end{aligned}$$

Führt man hier noch die Koordinate ϑ an Stelle der Zeit t vermöge der Beziehung

$$\text{(A.1.23)} \qquad \mathrm{d}\vartheta = \frac{v}{R}\,\mathrm{d}t$$

ein, so erhält man ersichtlich die Gln. (E.2) und (E.3) der Einleitung. In magnetischen Linsen mit $H_z = 0$ ist $H_z \varrho$ (im Nenner) natürlich nach Gl. (A.1.19) durch den Impuls auszudrücken ([35]).

R.Q.Twiss (unveröffentlicht) machte darauf aufmerksam, daß sich der Feldindex für Teilchen abweichenden Impulses ändert. Bezieht man die Überlegungen auf eine mit abgeändertem Impuls durchlaufene Bahn, so wird z.B. bei Gewinnung der für die z-Schwingung gültigen Gleichung durch ein etwas anderes, im Verhältnis der Impulse geändertes, magnetisches Feld dividiert. Die gleichzeitige Änderung des Krümmungsradius und der Bahnlänge ist bei hohem Feldindex vielfach zu vernachlässigen. Die Feldindizes n_{zz} und n_{zr} (Gl. (E.3.)) sind also zu multiplizieren mit p_0/p, wobei p_0 den Sollimpuls, p aber den wahren Impuls des Teilchens bedeutet. Wegen

$$\text{(A.1.24)} \qquad n_{rz} = n_{zr}\,, \qquad n_{rr} + n_{zz} = -\frac{R^2}{\varrho^2}$$

überträgt sich die Änderung des Feldindex auch auf die für die horizontale Schwingung maßgeblichen Größen. Wegen der dichten Lage der Resonanzpunkte darf dieser Effekt nicht übersehen werden.

b) Berücksichtigung der Beschleunigung. – Bei Berücksichtigung der Beschleunigung der Teilchen ist die Bewegungsgleichung (A.1.15) zu ersetzen durch

$$\text{(A.1.25)} \qquad (m\dot{\boldsymbol{R}})^{\cdot} = \frac{e}{c}\,\dot{\boldsymbol{R}} \times \boldsymbol{H} + e\boldsymbol{E}\,.$$

Die beschleunigende Spannung soll tangential zur Bahn ($\boldsymbol{E} = E\boldsymbol{t}_c$) wirken;

([35]) Diese Bewegungsgleichungen stellen noch nicht völlig den allgemeinsten Fall dar, da das magnetische Führungsfeld im Prinzip auch eine Komponente parallel zur Gleichgewichtsbahn besitzen kann. Nach freundlicher persönlicher Mitteilung durch Herrn E. D. Courant können die Bewegungsgleichungen aber auch dann, indem sie für ein mitgedrehtes Bezugssystem formuliert werden, in die oben angegebene Form gebracht werden.

es gilt dann für ein nicht ausgelenktes Teilchen

$$(m\dot{v})^{\cdot} = eE\,. \tag{A.1.26}$$

Bei Aufstellung der skalaren Bewegungsgleichungen ist jetzt einerseits der Term $\dot{m}\dot{\boldsymbol{R}}$ zu berücksichtigen, andererseits ist bei Bildung von $\ddot{\boldsymbol{R}}$ auch die Geschwindigkeit zu differenzieren. Somit erhält man unter Berücksichtigung von Gl. (A.1.26)

$$\left\{\begin{aligned} &(m\dot{\tau})^{\cdot} + \left(\frac{mvr}{\varrho}\right)^{\cdot} = 0 \\ &(m\dot{z})^{\cdot} = -\frac{ev}{c}\,(K_{32}z + K_{33}r) \\ &(m\dot{r})^{\cdot} - \frac{mv\dot{\tau}}{\varrho} = \frac{ev}{c}\,(K_{22}z + K_{23}r)\,. \end{aligned}\right. \tag{A.1.27}$$

Die Bedingung, daß die Geschwindigkeit der Teilchen in jedem Augenblick gleich der aus (A.1.26) folgenden Sollgeschwindigkeit ist, (d.h. daß wiederum Gl. (A.1.21), jetzt aber mit zeitlich veränderlichem v, gilt) erlaubt wiederum, die erste Gleichung fortzulassen, während sich die beiden verbleibenden Gleichungen jetzt folgendermaßen schreiben

$$\begin{aligned} (m\dot{z})^{\cdot} &= -\frac{ev}{c}\,(K_{32}z + K_{33}r)\,, \\ (m\dot{r})^{\cdot} &= \frac{ev}{c}\left(K_{22}z + \left(K_{33} + \frac{H_z}{\varrho}\right)r\right). \end{aligned} \tag{A.1.28}$$

Beide Gleichungen sind vom gleichen Typ. Es genügt daher die Diskussion etwa der ersten Gleichung. Die Zeit t soll wieder durch die Koordinate ϑ ausgedrückt werden, wobei jetzt die zeitliche Änderung der Geschwindigkeit v zu beachten ist. Man findet

$$(m\dot{z})^{\cdot} = \frac{mv^2}{R}\left(z'' + \frac{(mv)'}{mv}\,z'\right) \tag{A.1.29}$$

und somit

$$z'' + \frac{p'}{p}\,z' - (n_{zz}z + n_{zr}r) = 0\,; \tag{A.1.30}$$

vgl. Gl. (E.3). Es ist bemerkenswert, daß hier nur noch der Teilchenimpuls p, auftritt, nicht aber Masse und Geschwindigkeit einzeln. Einführung der Funktionen

$$z_0(\vartheta) = \sqrt{p}\,z(\vartheta)\,, \qquad r_0(\vartheta) = \sqrt{p}\,r(\vartheta) \tag{A.1.31}$$

verwandelt Gl. (A.1.30) in

$$z_0'' - \left\{\left(n_{zz} + \left(\frac{p'}{2p}\right)' - \left(\frac{p'}{2p}\right)^2\right)z_0 + n_{zr}r_0\right\} = 0\,. \tag{A.1.32}$$

Entsprechend gilt

$$r_0'' - \left\{ n_{rz} z_0 + \left(n_{rr} + \left(\frac{p'}{2p} \right)' - \left(\frac{p'}{2p} \right)^2 \right) r_0 \right\} = 0 \,. \tag{A.1.33}$$

Man hat also die Gleichungen einer gewöhnlichen Betatronschwingung ohne Beschleunigungsspannung, wobei die Feldindizes gegenüber den im stationären Problem auftretenden Größen allerdings etwas verändert sind. Der Einfluß dieser Zusatzterme auf die Gestalt der Bahn und auf die Größe μ ist für hinreichend langsam veränderliches p vernachlässigbar klein. In diesem Sinne kann hier von einer « adiabatischen Behandlung » gesprochen werden.

Würde sich der Feldindex $n(\vartheta)$ während des ganzen Beschleunigungszyklus nicht ändern, so könnte aus Gln. (A.1.31) und (A.1.32), (A.1.33) abgelesen werden, daß die Betatronschwingungen während des Beschleunigungszyklus streng umgekehrt proportional zur Quadratwurzel aus dem Teilchenimpuls gedämpft werden. Tatsächlich wird sich der Feldindex während des Beschleunigungszyklus ändern (unter anderem wegen Sättigung der Magnete). Aber auch dann wird der Übergang von den formalen Lösungen der Gln. (A.1.32), (A.1.33) zu den wirklichen Bahnen durch einen Faktor $p^{-\frac{1}{2}}$ vermittelt (Gl. (A.1.31)). Dieser Dämpfungsmechanismus wird als « adiabatische Dämpfung » bezeichnet und wurde wohl zuerst von KERST und SERBER ([36]) bemerkt.

ANHANG 2.

Zur störungsmässigen Behandlung kleiner Änderungen des Feldindex.

In Teil I, Abschn. 1*a*, wurde ein Verfahren der Störungsrechnung skizziert, das sich auf die Untersuchung des Einflusses kleiner Änderungen des Feldindex anwenden läßt. Die genauere Durchführung erfolgte in Teil I nur für die instabilen Bereiche, die infolge kleiner Fehler des Feldindex auftreten. Einige weitere Anwendungen und mathematische Ergänzungen sollen in diesem Anhang dargestellt werden. Unterabschn. *a* wird sich mit der Auswirkung von Änderungen des Feldindex im Innern eines stabilen Gebiets befassen, wobei im Gegensatz zu der Anwendung im Hauptteil der Arbeit angenommen werden soll, daß das ungestörte Q nicht gerade ganz- oder halbzahlig ist. Unterabschn. *b* wird die Auswirkung solcher Änderungen an der Grenze eines stabilen Bereichs behandeln, wobei entweder an eine Bereichsgrenze des ungestörten Beschleunigers (μ ein ganzzahliges Vielfaches von π) oder an die Grenze eines instabilen Bereichs infolge Fehlern des Feldindex gedacht werden kann. Unterabschn. *c* wird mathematische Ergänzungen zu der im Hauptteil dargestellten Störungsrechnung zur Gewinnung instabiler Bereiche bringen.

a) Das Innere eines stabilen Gebiets. – Die Rechnungen dieses und des folgenden Unterabschnitts lassen sich sowohl auf einen vollen Umlauf als auch

([36]) D. W. KERST und R. SERBER: *Phys. Rev.*, **60**, 53 (1941).

—bei streng periodischem Feldindex und streng periodischer Änderung — auf einen einzigen Periodizitätsbereich anwenden. Im zweiten Fall ist statt Gl. (I.2.1) zu fordern

$$\text{(A.2.1)} \qquad \tilde{r}_1(\vartheta+\vartheta_0)=\exp[i\tilde{\mu}]\,\tilde{r}_1(\vartheta)\,, \qquad \tilde{r}_2(\vartheta+\vartheta_0)=\exp[-i\tilde{\mu}]\,\tilde{r}_2(\vartheta)$$

und statt Gl. (I.2.2) zu setzen

$$\text{(A.2.2)} \qquad \begin{cases} r_1(\vartheta)=v_1(\vartheta)\exp[i\mu\vartheta/\vartheta_0]\,, & \tilde{r}_1(\vartheta)=\tilde{v}_1(\vartheta)\exp[i\tilde{\mu}\vartheta/\vartheta_0]\,, \\ r_2(\vartheta)=v_2(\vartheta)\exp[-i\mu\vartheta/\vartheta_0]\,, & \tilde{r}_2(\vartheta)=\tilde{v}_2(\vartheta)\exp[-i\tilde{\mu}\vartheta/\vartheta_0]\,. \end{cases}$$

Um die nachfolgenden Überlegungen genügend allgemein zu halten, soll in den Gln. (I.2.3) bis (I.2.10) statt Q bzw. $\tilde{Q}$ die Bezeichnung q bzw. $\tilde{q}$ eingeführt werden. Bei Anwendung auf einen einzigen Periodizitätsbereich ist dann zu setzen

$$\text{(A.2.3)} \qquad q=\frac{\mu}{\vartheta_0}\,, \quad \tilde{q}=\frac{\tilde{\mu}}{\vartheta_0}\,.$$

Ferner soll das jeweilige Periodizitätsintervall (ϑ_0 bzw. 2π) allgemein mit θ_0 bezeichnet werden.

Die zweite Gl. (I.2.8) oder

$$\text{(A.2.4)} \qquad \tilde{v}_1^{(1)\prime\prime}(\vartheta)+2iq\tilde{v}_1^{(1)\prime}(\vartheta)-(n(\vartheta)+q^2)\tilde{v}_1^{(1)}(\vartheta)= \\ =-2i\tilde{q}^{(1)}v_1'(\vartheta)+2\tilde{q}^{(1)}qv_1(\vartheta)+\delta n(\vartheta)v_1(\vartheta)$$

kann in geschlossener Form nach $\tilde{v}_1^{(1)}(\vartheta)$ aufgelöst werden. Bezeichnet man die rechte Seite für den Augenblick durch $f(\vartheta)$, so findet man in leichter Verallgemeinerung der Gl. (I.1.2)

$$\text{(A.2.5)} \qquad \tilde{v}_1^{(1)}(\vartheta)=a_{11}^{(1)}w_1(\vartheta)+a_{12}^{(1)}w_2(\vartheta)+ \\ +\int_0^{\vartheta}\mathrm{d}\vartheta'\,(W(\vartheta'))^{-1}f(\vartheta')\{w_1(\vartheta')w_2(\vartheta)-w_2(\vartheta')w_1(\vartheta)\}\,,$$

wobei für die aus zwei linear unabhängigen Lösungen $w_1(\vartheta)$, $w_2(\vartheta)$ der homogenen Gleichung gebildete Wronskideterminante gilt

$$\text{(A.2.6)} \qquad W(\vartheta)=W(0)\exp[-2iq\vartheta]\,.$$

Zweckmäßig wählt man

$$\text{(A.2.7)} \qquad w_1(\vartheta)=v_1(\vartheta)\,, \qquad w_2(\vartheta)=v_2(\vartheta)\exp[-2iq\vartheta]\,;$$

die aus $w_1(\vartheta)$, $w_2(\vartheta)$ gebildete Wronskideterminante ist dann für $\vartheta=0$ mit der Wronskideterminante (E.9) identisch und soll deshalb kurz mit W be-

zeichnet werden. Gl. (A.2.5) schreibt sich somit

$$\text{(A.2.8)}\quad \tilde{v}_1^{(1)}(\vartheta) = a_{11}^{(1)} v_1(\vartheta) + a_{12}^{(1)} v_2(\vartheta) \exp[-2iq\vartheta] + \\ + \frac{1}{W}\int_0^{\vartheta} \mathrm{d}\vartheta' \exp[2iq\vartheta']\{-2i\tilde{q}^{(1)} v_1'(\vartheta') + 2\tilde{q}^{(1)} q v_1(\vartheta') + \delta n(\vartheta') v_1(\vartheta')\}\cdot \\ \cdot\{v_1(\vartheta') v_2(\vartheta) \exp[-2iq\vartheta] - v_2(\vartheta') v_1(\vartheta) \exp[-2iq\vartheta']\}\,.$$

Jetzt ist die Periodizitätsforderung (I.2.9) — für θ_0 statt 2π — anzuwenden, aus der nach entsprechenden Überlegungen wie in Zusammenhang mit Gl. (I.2.13) einerseits folgt

$$\text{(A.2.9)}\quad \int_0^{\theta_0} \mathrm{d}\vartheta' \{-2i\tilde{q}^{(1)} v_1'(\vartheta') + 2\tilde{q}^{(1)} q v_1(\vartheta') + \delta n(\vartheta') v_1(\vartheta')\} v_2(\vartheta') = 0\,,$$

und andererseits für den Koeffizienten $a_{12}^{(1)}$

$$\text{(A.2.10)}\quad a_{12}^{(1)} = \frac{\int_0^{\theta_0} \mathrm{d}\vartheta' \{-2i\tilde{q}^{(1)} v_1'(\vartheta') + 2\tilde{q}^{(1)} q v_1(\vartheta') + \delta n(\vartheta') v_1(\vartheta')\} v_1(\vartheta') \exp[2iq\vartheta']}{W(\exp[2iq\theta_0] - 1)}\,.$$

Der Koeffizient $a_{11}^{(1)}$ bleibt unbestimmt; das rührt daher, daß man jede Lösung des gestörten Problems mit einem konstanten Faktor multiplizieren kann, der bei verschwindender Störung gegen eins geht. Ein analoges Verhalten ist von den höheren Näherungen der Schrödingerschen Störungsrechnung der Quantentheorie wohlbekannt. Es sei noch ausdrücklich angemerkt, daß Gl. (A.2.10) nur sinnvoll ist, wenn $\exp[2iq\theta_0]$ von $+1$ verschieden ist; es darf also keine « Entartung » im Sinne von Teil I, Abschn. 2*a* vorliegen.

Aus Gl. (A.2.9) ist die Änderung $\tilde{q}^{(1)}$ der Größe q zu bestimmen. Anwendung der Periodizitätseigenschaften von $v_1(\vartheta)$, $v_2(\vartheta)$ und der Definition der Wronskideterminante liefert zunächst

$$\text{(A.2.11)}\quad \int_0^{\theta_0} (v_1'(\vartheta) + iqv_1(\vartheta)) v_2(\vartheta) \mathrm{d}\vartheta = -\frac{W\theta_0}{2}$$

(ein Spezialfall ist die zweite Gl. (I.2.17)) und somit

$$\text{(A.2.12)}\quad \tilde{q}^{(1)} = -\frac{1}{iW\theta_0}\int_0^{\theta_0} \delta n(\vartheta) v_1(\vartheta) v_2(\vartheta)\,\mathrm{d}\vartheta = -\frac{1}{iW\theta_0}\int_0^{\theta_0} \delta n(\vartheta) r_1(\vartheta) r_2(\vartheta)\,\mathrm{d}\vartheta\,.$$

Sieht man $\delta n(\vartheta)$ hier insbesondere als eine streng periodische Änderung des idealen Feldindex an, sei also $\delta n(\vartheta)$ periodisch mit ϑ_0, so kann von Gl. (A.2.3) Gebrauch gemacht werden und man erhält

$$\text{(A.2.13)}\quad \tilde{\mu}^{(1)} = -\frac{1}{iW}\int_0^{\vartheta_0} \delta n(\vartheta) r_1(\vartheta) r_2(\vartheta)\,\mathrm{d}\vartheta\,.$$

Die Änderung der Zahl der Betatronwellenlängen auf dem Umfang ergibt sich in dieser Näherung wegen Gl. (E.12) dann zu

$$\tilde{Q}^{(1)} = -\frac{1}{2\pi i W}\int_0^{2\pi} \delta n(\vartheta) r_1(\vartheta) r_2(\vartheta)\,\mathrm{d}\vartheta \tag{A.2.14}$$

in voller Übereinstimmung mit Gl. (I.2.24). Diese Gleichung war jedoch damals speziell für ganz- oder halbzahliges Q abgeleitet worden. Übrigens erlaubt auch Gl. (A.2.10) eine vereinfachende Umformung. Es ergibt sich

$$a_{12}^{(1)} = \frac{\tilde{q}^{(1)}(r_1(0))^2}{iW} + \frac{\int_0^{\vartheta_0} \delta n(\vartheta)(r_1(\vartheta))^2\,\mathrm{d}\vartheta}{W(\exp[2i\vartheta_0]-1)}\,. \tag{A.2.15}$$

Aus Gl. (A.2.13) kann eine Aussage über das Vorzeichen der bereits als imaginär erkannten [37] Wronskideterminante gewonnen werden. Zu diesem Zweck werde

$$\delta n(\vartheta) = \delta k \cdot n(\vartheta) \tag{A.2.16}$$

gesetzt, also $n(\vartheta)$ überall mit dem gleichen nur wenig von eins verschiedenen Faktor multipliziert. Das dann entstehende Integral kann unter Benutzung der Gl. (E.6) für die freien Betatronschwingungen und durch nachfolgende partielle Integration umgeformt werden

$$\frac{\tilde{\mu}^{(1)}}{\delta k} = n\frac{\mathrm{d}\mu}{\mathrm{d}n} = -\frac{1}{iW}\int_0^{\vartheta_0} n(\vartheta) r_1(\vartheta) r_2(\vartheta)\,\mathrm{d}\vartheta = \frac{1}{iW}\int_0^{\vartheta_0} r_1'(\vartheta) r_2'(\vartheta)\,\mathrm{d}\vartheta\,. \tag{A.2.17}$$

Bei der Umformung muß beachtet werden, daß das Produkt $r_1'(\vartheta)r_2(\vartheta)$ in ϑ_0 periodisch ist. Wegen $r_2(\vartheta) = r_1^*(\vartheta)$ ist nun das Integral $\int_0^{\vartheta_0} r_1'(\vartheta) r_2'(\vartheta)\,\mathrm{d}\vartheta$ sicher positiv. Somit gewinnt man folgende Aussage: Die aus den Floquetschen Lösungen $r_1(\vartheta)$ und $r_2(\vartheta)$ gebildete Wronskideterminante W stellt eine mit der imaginären Einheit multiplizierte positive Größe dar, wenn die für $r_1(\vartheta)$ nach Gln. (E.10) und (E.11) definierte Größe μ bei gleichmäßiger Vergrößerung des Feldindex ($\delta k > 0$) abnimmt. Derartige Vorzeichenüberlegungen spielen eine Rolle bei Untersuchung derjenigen instabilen Bereiche, die durch eine Verwindung der Sektoren gegen einander hervorgerufen werden. Übrigens erkennt man aus den vorgetragenen Überlegungen auch, daß sich μ bei einer solchen monotonen Änderung des Feldindex innerhalb eines Stabilitätsgebiets selbst monoton ändert.

Die Berechnung höherer Näherungen verläuft in genau der gleichen Weise. Die Periodizitätsbedingung verhilft einerseits (mittels der für den Faktor von $c_1(\vartheta)$ geltenden Gleichung) zur entsprechenden Näherung von $\tilde{q}$, während aus der anderen Gleichung eine Aussage über den einen Entwicklungskoeffizienten

---

[37] Vgl. Teil I, Bemerkungen vor Gl. (I.1.4).

gewonnen wird. Auch in den höheren Näherungen bleibt natürlich eine Unbestimmtheit der Koeffizienten. Es ist zu vermuten, daß die entstehende Reihe für kleine Änderungen $\delta n(\vartheta)$ konvergiert, wobei der Konvergenzbereich aber nie bis an die Grenze des nächsten instabilen Bereichs heranreichen wird.

b) *Die Grenze eines stabilen Gebiets.* – An der Grenze eines stabilen Gebiets soll wieder — wie in Teil I, Abschn. $2a$ — q zu null umdefiniert und die Periodizitätsbedingung für $r_j(\vartheta)$, $\tilde{v}_j(\vartheta)$ nötigenfalls durch eine Antiperiodizitätsbedingung ersetzt werden.

Es sei zunächst angenommen, daß an der Grenze des Gebiets alle Lösungen periodisch bzw. antiperiodisch sind. Es wird sich zeigen, daß dann bei einer gleichmäßigen Änderung des Feldindex nach Gl. (A.2.16) keine Grenze eines stabilen Gebiets vorliegt. Tritt daher bei einer solchen Änderung eine echte Grenze des Stabilitätsgebietes auf, so kann es dort nur eine einzige periodische bzw. antiperiodische Lösung geben, während für alle anderen Lösungen eine Gleichung vom Typ der Gl. (E.14) gilt. Damit ist dann eine am Schluß der Einleitung aufgestellte Behauptung bewiesen.

Es sei also vorausgesetzt, daß alle Lösungen an der Gebietsgrenze periodisch (antiperiodisch) sind. Dann lassen sich die Rechnungen von Teil I, Abschn. $2a$ ohne Änderung übernehmen. Gemäß Gl. (I.2.19) findet man

$$\text{(A.2.18)} \qquad \tilde{q}^{(1)2} = \frac{(\int_0^{\theta_0}\delta n(\vartheta) r_1(\vartheta) r_2(\vartheta)\,\mathrm{d}\vartheta)^2 - |\int_0^{\theta_0}\delta n(\vartheta)(r_1(\vartheta))^2\,\mathrm{d}\vartheta|^2}{(2\pi i W)^2} .$$

Entsprechende Umformungen wie in Gl. (A.2.17) zeigen nun, daß der Zähler für $\delta n(\vartheta)$ gemäß Gl. (A.2.16) nach der Cauchy-Schwarzschen Ungleichung nie negativ ist. Deshalb ist $\tilde{q}^{(1)}$ reell und man hat bei kleinen Änderungen δk stets Stabilität; es liegt bei dieser Änderung des Feldindex also keine Grenze des stabilen Gebiets vor.

Nunmehr sei angenommen, daß nur eine einzige periodische (antiperiodische) Lösung an der Bereichsgrenze existiert. Neben dieser Lösung werde eine zweite, $r_0(\vartheta)$, verwendet. Gemäß Gln. (E.13) und (E.14) gilt

$$\text{(A.2.19)} \qquad r_{\mathrm{per}}(\vartheta + \theta_0) = \pm r_{\mathrm{per}}(\vartheta) , \qquad r_0(\vartheta + \theta_0) = \pm (r_0(\vartheta) + a r_{\mathrm{per}}(\vartheta)) .$$

Die Größe W bedeute die aus diesen beiden Lösungen gebildete Wronskideterminante. Man überzeugt sich leicht, daß der Quotient a/W von der speziellen Wahl der zweiten Lösung $r_0(\vartheta)$ unabhängig ist. Setzt man jetzt die Störungsrechnung nach demselben Schema wie vorher an, so gelangt man im allgemeinen (falls nicht eine spezielle Bedingung für $\delta n(\vartheta)$ erfüllt ist) zu einem Widerspruch. Es ist also eine andere Größenordnung in der Entwicklung von $\tilde{q}$ anzunehmen. Die zweite und dritte Gleichung des in Gl. (I.2.8) angegebenen Störungsschemas soll jetzt lauten

$$\text{(A.2.20)} \qquad \begin{cases} \tilde{v}_1^{(1)\prime}(\vartheta) - n(\vartheta)\tilde{v}_1^{(1)}(\vartheta) = -2i\tilde{q}^{(1)} r'_{\mathrm{per}}(\vartheta) , \\ \tilde{v}_1^{(2)\prime}(\vartheta) - n(\vartheta)\tilde{v}_1^{(2)}(\vartheta) = -2i\tilde{q}^{(2)} r'_{\mathrm{per}}(\vartheta) - 2i\tilde{q}^{(1)}\tilde{v}_1^{(1)\prime}(\vartheta) + \\ \qquad\qquad + \tilde{q}^{(1)2} r_{\mathrm{per}}(\vartheta) + \delta n(\vartheta) r_{\mathrm{per}}(\vartheta) . \end{cases}$$

Die Lösung der ersten Gleichung lautet

$$\text{(A.2.21)}\quad \tilde{v}_1^{(1)}(\vartheta)=a_{11}^{(1)} r_{\mathrm{per}}(\vartheta)+a_{12}^{(1)} r_0(\vartheta)-\frac{2i\tilde{q}^{(1)}}{W}\int_0^{\vartheta} \mathrm{d}\vartheta'\, r'_{\mathrm{per}}(\vartheta')\left(r_{\mathrm{per}}(\vartheta')\, r_0(\vartheta)-r_0(\vartheta')\, r_{\mathrm{per}}(\vartheta)\right).$$

Man entnimmt aus der Periodizitätsforderung

$$\text{(A.2.22)}\quad a_{12}^{(1)}=-\frac{2i\tilde{q}^{(1)}}{aW}\int_0^{\theta_0} r'_{\mathrm{per}}(\vartheta)\, r_0(\vartheta)\,\mathrm{d}\vartheta\,,$$

wobei die Größe $\tilde{q}^{(1)}$ in dieser Näherung noch unbestimmt bleibt und sich erst im nächsten Schritt ergibt. Aus der Periodizitätsforderung folgt jetzt keine zweite Gleichung, da gilt

$$\text{(A.2.23)}\quad \int_0^{\theta_0} r'_{\mathrm{per}}(\vartheta)\, r_{\mathrm{per}}(\vartheta)\,\mathrm{d}\vartheta=0\,.$$

Der Koeffizient $a_{11}^{(1)}$ bleibt natürlich wieder unbestimmt. Einfache Umformungen liefern für $\tilde{v}_1^{(1)}(\vartheta)$ den geschlossenen Ausdruck (bei spezieller Wahl von $a_{11}^{(1)}$)

$$\text{(A.2.24)}\quad \tilde{v}_1^{(1)}(\vartheta)=i\tilde{q}^{(1)}\left(\frac{\theta_0 r_0(\vartheta)}{a}-\vartheta r_{\mathrm{per}}(\vartheta)\right).$$

Aus der zweiten Gl. (A.2.20) erhält man

$$\begin{aligned}\text{(A.2.25)}\quad \tilde{v}_1^{(2)}(\vartheta)&=a_{11}^{(2)} r_{\mathrm{per}}(\vartheta)+a_{12}^{(2)} r_0(\vartheta)+\\&+\frac{1}{W}\int_0^{\vartheta}\mathrm{d}\vartheta'\left(-2i\tilde{q}^{(2)} r'_{\mathrm{per}}(\vartheta')-2i\tilde{q}^{(1)}\tilde{v}_1^{(1)\prime}(\vartheta')-\tilde{q}^{(1)2} r_{\mathrm{per}}(\vartheta')+\delta n(\vartheta')\, r_{\mathrm{per}}(\vartheta')\right)\cdot\\&\cdot\left(r_{\mathrm{per}}(\vartheta')\, r_0(\vartheta)-r_0(\vartheta')\, r_{\mathrm{per}}(\vartheta)\right),\end{aligned}$$

und gewinnt aus der Periodizitätsbedingung die beiden Gleichungen

$$\text{(A.2.26)}\quad \begin{cases} 0=\displaystyle\int_0^{\theta_0}\Big(-2i\tilde{q}^{(2)} r'_{\mathrm{per}}(\vartheta')-2i\tilde{q}^{(1)}\tilde{v}_1^{(1)\prime}(\vartheta')+\tilde{q}^{(1)2} r_{\mathrm{per}}(\vartheta')+\delta n(\vartheta')\, r_{\mathrm{per}}(\vartheta')\Big)\cdot \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\cdot r_{\mathrm{per}}(\vartheta')\,\mathrm{d}\vartheta', \\ a_{12}^{(2)}=-\dfrac{1}{aW}\displaystyle\int_0^{\theta_0}\Big(-2i\tilde{q}^{(2)} r'_{\mathrm{per}}(\vartheta')-2i\tilde{q}^{(1)}\tilde{v}_1^{(1)\prime}(\vartheta')+\tilde{q}^{(1)2} r_{\mathrm{per}}(\vartheta')+ \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad+\delta n(\vartheta')\, r_{\mathrm{per}}(\vartheta')\Big)\, r_0(\vartheta')\,\mathrm{d}\vartheta'\,. \end{cases}$$

Die erste Gleichung erlaubt die Bestimmung von $\tilde{q}^{(1)}$. Nach Einsetzen von Gl. (A.2.24) erhält man unter Benutzung von Gl. (A.2.23) nach einiger Zwi-

schenrechnung

$$(\theta_0\tilde{q}^{(1)})^2 = -\frac{a}{W}\int_0^{\vartheta_0}\delta n(\vartheta)(r_{\mathrm{per}}(\vartheta))^2\,\mathrm{d}\vartheta, \tag{A.2.27}$$

wobei a/W, wie oben bemerkt, von der speziellen Wahl der Funktion $r_0(\vartheta)$ nicht abhängt. Eine weitere Auswertung von Gl. (A.2.27) und die Berechnung höherer Näherungen soll hier nicht erfolgen.

c) *Instabile Bereiche infolge Fehlern im Feldindex.* – Es sollen einige mathematische Ergänzungen nachgetragen werden zu den Rechnungen, die in Teil I, Abschn. 2 zur Ermittlung der instabilen Bereiche durchgeführt wurden. Es soll nämlich gezeigt werden, daß sich die noch unbestimmt gebliebenen Koeffizienten $a_{11}^{(1)}$ und $a_{12}^{(1)}$ (vgl. Gl. (I.2.13)) und die zweite Näherung von $\tilde{Q}$ wirklich aus der zweiten störungstheoretischen Näherung gewinnen lassen. Hierfür sei $\tilde{Q}^{(1)} \neq 0$ vorausgesetzt. Die Gleichung zur Bestimmung von $\tilde{r}_1^{(2)}(\vartheta)$ lautet

$$\tilde{v}_1^{(2)\prime}(\vartheta) - n(\vartheta)\tilde{r}_1^{(2)}(\vartheta) = -2i\tilde{Q}^{(1)}\tilde{v}_1^{(1)\prime}(\vartheta) + \delta n(\vartheta)\tilde{v}_1^{(1)}(\vartheta) - \\ - 2i\tilde{Q}^{(2)}\tilde{v}_1^{(0)\prime}(\vartheta) + \tilde{Q}^{(1)^2}\tilde{v}_1^{(0)}(\vartheta)\,. \tag{A.2.28}$$

Es ist bemerkenswert, daß die Funktion $\tilde{r}_1^{(1)}(\vartheta)$, deren Koeffizienten $a_{11}^{(1)}$ und $a_{12}^{(1)}$ ja bestimmt werden sollen, auf der rechten Seite mit denselben Faktoren auftritt wie die Funktion $\tilde{r}_1^{(0)}(\vartheta)$ in Gl. (I.2.8) (mit $\tilde{Q}^{(0)} = 0$). Man sieht leicht allgemein, daß die Funktion $\tilde{r}_1^{(n)}(\vartheta)$ in der Gleichung für die $(n+1)$. Näherung stets in der Form $-2i\tilde{Q}^{(1)}r_1^{(n)\prime}(\vartheta) + \delta n(\vartheta)r_1^{(n)}(\vartheta)$ auftritt; die Ermittlung der Koeffizienten $a_{11}^{(n)}$, $a_{12}^{(n)}$ erfolgt daher auch in allen weiteren Schritten nach demselben Schema. Der in der jeweiligen Näherung zu bestimmende Beitrag $\tilde{Q}^{(n+1)}$ zu $\tilde{Q}$ tritt dagegen stets in der Form $-2i\tilde{Q}^{(n+1)}\tilde{v}_1^{(0)\prime}(\vartheta)$ auf.

Auf der rechten Seite der Gl. (A.2.28) sollen jetzt $\tilde{v}_1^{(0)}(\vartheta)$ und $\tilde{r}_1^{(1)}(\vartheta)$ aus Gln. (I.2.12), (I.2.13) eingesetzt werden, wobei nur derjenige Teil, der $\tilde{Q}^{(2)}$ oder die Koeffizienten $a_{11}^{(1)}$, $a_{12}^{(1)}$ enthält, explizit ausgeschrieben sei

$$\tilde{v}_1^{(2)\prime}(\vartheta) - n(\vartheta)\tilde{r}_1^{(2)}(\vartheta) = -2i\tilde{Q}^{(1)}(a_{11}^{(1)}r_1'(\vartheta) + a_{12}^{(1)}r_2'(\vartheta)) + \\ + \delta n(\vartheta)(a_{11}^{(1)}r_1(\vartheta) + a_{12}^{(1)}r_2(\vartheta)) - 2i\tilde{Q}^{(2)}(a_{11}^{(0)}r_1'(\vartheta) + a_{12}^{(0)}r_2'(\vartheta)) + \dots\,. \tag{A.2.29}$$

Als Lösung erhält man

$$\tilde{r}_1^{(2)}(\vartheta) = a_{11}^{(2)}r_1(\vartheta) + a_{12}^{(2)}r_2(\vartheta) - \frac{1}{W}\int_0^{\vartheta}\mathrm{d}\vartheta'[-2i\tilde{Q}^{(1)}(a_{11}^{(1)}r_1'(\vartheta') + a_{12}^{(1)}r_2'(\vartheta')) + \\ + \delta n(\vartheta')(a_{11}^{(1)}r_1(\vartheta') + a_{12}^{(1)}r_2(\vartheta')) - 2i\tilde{Q}^{(2)}(a_{11}^{(0)}r_1'(\vartheta') + a_{12}^{(0)}r_2'(\vartheta')) + \dots] \cdot \\ \cdot (r_1(\vartheta')r_2(\vartheta) - r_2(\vartheta')r_1(\vartheta))\,; \tag{A.2.30}$$

die Koeffizienten $a_{11}^{(2)}$, $a_{12}^{(2)}$ sind aus der dritten Näherung zu bestimmen. Die

Periodizitätsbedingung (I.2.9) liefert die Gleichungen (benutzt wurde Gl. (I.2.17))

$$\text{(A.2.31)} \quad \begin{cases} a_{11}^{(1)} N_{11} + a_{12}^{(1)} (N_{21} - 2\pi i W \tilde{Q}^{(1)}) = 2\pi i W \tilde{Q}^{(2)} a_{12}^{(0)} + \dots , \\ a_{11}^{(1)} (N_{12} + 2\pi i W \tilde{Q}^{(1)}) + a_{12}^{(1)} N_{22} = - 2\pi i W \tilde{Q}^{(2)} a_{11}^{(0)} + \dots . \end{cases}$$

Wiederum sind hier nur diejenigen Teile, die $\tilde{Q}^{(2)}$ oder $a_{11}^{(1)}$, $a_{12}^{(1)}$ enthalten, explizit angegeben. Die Koeffizienten des homogenen Teils dieses Gleichungssystems sind identisch mit denen, die im Gleichungssystem (I.2.18) für $a_{11}^{(0)}$, $a_{12}^{(0)}$ auftraten. Ein entsprechender Sachverhalt wiederholt sich in allen höheren Näherungen. Hieraus folgt zunächst, daß man zu einer Lösung $a_{11}^{(1)}$, $a_{12}^{(1)}$ der Gl. (A.2.31) ein beliebiges Vielfaches von $a_{11}^{(0)}$, $a_{12}^{(0)}$ hinzuaddieren kann; das ist wieder die Unbestimmtheit in den Koeffizienten der höheren Näherungen, die uns schon vielfach begegnet ist und deren Grund in Unterabschn. *a* angegeben wurde.

Die aus den Koeffizienten des homogenen Bestandteils von Gl. (A.2.31) gebildete Determinante verschwindet; dies war in der ersten Näherung gerade die Bedingung zur Bestimmung von $\tilde{Q}^{(1)}$. Die beiden Gln. (A.2.31) sind daher im allgemeinen gar nicht mit einander verträglich. Aus der Bedingung der Verträglichkeit läßt sich $\tilde{Q}^{(2)}$ bestimmen. Ohne die Berechnung im einzelnen durchzuführen, soll gezeigt werden, daß man auf diese Weise wirklich zu einem eindeutigen Wert von $\tilde{Q}^{(2)}$ gelangt. Da $\tilde{Q}^{(n)}$ in der n. Näherung in der gleichen Weise auftritt wie $\tilde{Q}^{(2)}$ in der vorliegenden zweiten Näherung, ist damit gezeigt, daß sich auch in höheren Näherungen keine neuen Schwierigkeiten ergeben.

Man bezeichne die rechten Seiten des Gleichungssystems (A.2.31) mit B_1, B_2, also

$$\text{(A.2.32)} \qquad B_1 = 2\pi i W \tilde{Q}^{(2)} a_{12}^{(0)} + \dots , \qquad B_2 = - 2\pi i W \tilde{Q}^{(2)} a_{11}^{(0)} + \dots .$$

Die Lösbarkeitsbedingung für Gl. (A.2.31) lautet dann

$$\text{(A.2.33)} \qquad \begin{cases} B_1 (N_{12} + 2\pi i W \tilde{Q}^{(1)}) - B_2 N_{11} = 0 , \\ B_1 N_{22} - B_2 (N_{21} - 2\pi i W \tilde{Q}^{(1)}) = 0 . \end{cases}$$

Die beiden Gleichungen sind linear abhängig, eine von ihnen kann jedoch unter Umständen identisch verschwinden. Es sei etwa angenommen, daß die Größen N_{22} und $N_{21} - 2\pi i W \tilde{Q}^{(1)}$ nicht beide gleich null sind; es ist dann auch $a_{11}^{(0)} \neq 0$. Jetzt ist die zweite Bedingungsgleichung (A.2.33) heranzuziehen

$$\text{(A.2.34)} \quad 0 = B_1 N_{22} - B_2 (N_{21} - 2\pi i W \tilde{Q}^{(1)}) = \\ = 2\pi i W \tilde{Q}^{(2)} (a_{12}^{(0)} N_{22} + a_{11}^{(0)} (N_{21} - 2\pi i W \tilde{Q}^{(1)})) + \dots .$$

Unter Benutzung der ersten Gl. (I.2.18) und Beachtung von $N_{21} = N_{12}$ erhält man somit

$$\text{(A.2.35)} \qquad 0 = - 2(2\pi i W)^2 \tilde{Q}^{(2)} \tilde{Q}^{(1)} a_{11}^{(0)} + \dots ,$$

wobei die fortgelassenen, durch Punkte angedeuteten Glieder wohlbestimmte Ausdrücke darstellen, die $\tilde{Q}^{(2)}$, $a_{11}^{(1)}$, $a_{12}^{(1)}$ nicht mehr enthalten. Wegen der Voraussetzungen $\tilde{Q}^{(1)} \neq 0$, $a_{11}^{(0)} \neq 0$ erlaubt Gl. (A.2.35) die eindeutige Auflösung nach $\tilde{Q}^{(2)}$. Koeffizienten $a_{11}^{(1)}$, $a_{12}^{(1)}$ können anschließend aus einer der Gln. (A.2.31) gewonnen werden.

Die so erhaltene Reihenentwicklung versagt in unmittelbarer Nähe der Grenze des instabilen Bereichs. Möchte man die Lage dieser Bereichsgrenze, über Gl. (I.2.28) hinaus in höheren Näherungen berechnen, so ist in einer wesentlich anderen Weise vorzugehen. Jedoch sei hierauf nicht eingegangen.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

Some Remarks on the Asymptotic Behaviour of the Cosmic Ray Cascade for Large Depth of the Absorber.

I.

Estimation of the Factorial Moments.

K. URBANIK

Institute of Mathematics of the Polish Academy of Science - Wrocław

(ricevuto il 12 Aprile 1955)

CONTENTS: 1. Formulation of the Problem. – 2. The Proof.

---

1. – Formulation of the Problem.

It is well known that the factorial moments can be expressed by means of the Mellin transform (1) e.g. the first moment

$$S_1(\varepsilon, x) = e^{-x} \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \mathrm{d}S \, \frac{\varepsilon^{-s}}{S} \exp\left[w(S)x\right]. \tag{1}$$

But this integral form of solution is not suitable for some considerations.

In our paper we investigate the behaviour of the moments for large depth of the absorber. We take advantage of the equations satisfied by these moments. No use, however, is made of the expressions, like (1).

We confine ourselves to the case of nucleon cascade in homogeneous nuclear matter. The results may be generalized to the case of nucleon cascade in finite absorber and to electron-photon cascade.

Let us denote the k-th factorial moment in the cascade theory formalism

(1) See e.g. H. MESSEL: *Proc. Roy. Irish Acad.*, A 54, 125 (1951); *Proc. Phys. Soc.*, A 64, 726 (1951), A 65, 465 (1952).

of JÁNOSSY (2) by $S_k(\varepsilon, x)$, i.e.

$$S_k(\varepsilon, x) = \sum_{n=k}^{[\varepsilon^{-1}]} n(n-1) \dots (n-k+1) P(n \mid \varepsilon, x) \tag{2}$$

where $P(n \mid \varepsilon, x)$ denotes the probability that at depth x there are n nucleons of energy $> \varepsilon$ due to one primary nucleon of energy 1.

We present the solution of the question suggested to us by ŁOPUSZAŃSKI. We have proved that for $0 < \varepsilon < 1$ and $k = 1, 2, 3, \dots [\varepsilon^{-1}]$

$$S_k(\varepsilon, x) = e^{-kx} W_k(\varepsilon, x)\,, \tag{3a}$$

where for each $\beta > 0$

$$\lim_{x \to \infty} \{e^{-\beta x} W_k(\varepsilon, x)\} = 0 \tag{3b}$$

and

$$\lim_{x \to \infty} W_k(\varepsilon, x) = \infty\,. \tag{3c}$$

The applications of (3) are given by ŁOPUSZAŃSKI in two subsequent papers (3).

2. – The Proof.

In order to demonstrate (3) it is sufficient to show that for $0 < \varepsilon < 1$ and $k = 1, 2, \dots [\varepsilon^{-1}]$

$$\lim_{x \to \infty} e^{\lambda x} S_k(\varepsilon, x) = \begin{cases} \infty & \text{for } \lambda = 1 \quad (4a) \\ 0 & \text{for } \lambda < 1\,. \quad (4b) \end{cases}$$

We get from the integral G-Equation of JÁNOSSY (2) the integral equation satisfied by the factorial moments for $0 < \varepsilon < 1$

$$S_k(\varepsilon, x) = \int_0^x e^{-(x-\xi)} \int_0^1 \int_0^1 w(\varepsilon', \varepsilon'') \cdot \left\{ \sum_{n=0}^{k} \binom{k}{n} S_n\left(\frac{\varepsilon}{\varepsilon'}, \xi\right) S_{k-n}\left(\frac{\varepsilon}{\varepsilon''}, \xi\right) \right\} \mathrm{d}\varepsilon' \, \mathrm{d}\varepsilon'' \, \mathrm{d}\xi + \delta_1^k e^{-x}\,, \tag{5}$$

where $S_0(\varepsilon, \xi) \equiv 1$.

---

(2) L. JÁNOSSY: *Proc. Phys. Soc.*, A **63**, 241 (1950); A **66**, 117 (1953).

(3) J. ŁOPUSZAŃSKI: see this issue, pag. 1150, 1161.

To prove (4*a*) we multiply both sides of (5) by e^{x}: consequently the right-hand side of (5) depends on x only by means of the upper limit of integration. In this equation the subintegral functions are positive and therefore it follows that $W_k(\varepsilon, x)$ is an increasing function of increasing x. It follows from (5) that

$$W_k(\varepsilon, x) \geqslant \int_0^1 \int_0^1 w(\varepsilon', \varepsilon'')\,\mathrm{d}\varepsilon'' \cdot \int_0^x W_k\left(\frac{\varepsilon}{\varepsilon'}, \xi\right)\mathrm{d}\xi\,\mathrm{d}\varepsilon' ,$$

and the right-hand side of this unequality tends to infinity, when x tends to infinity.

From the definition (2) we get

$$S_k(\varepsilon, x) \leqslant [\varepsilon^{-1}]^k S_1(\varepsilon, x)$$

consequently it is sufficient to show (4*b*) for $k=1$ only.

Let us perform a transformation of $S_1(\varepsilon, x)$

$$M(x) = e^{\lambda x} \int_0^1 S_1(\varepsilon, x)\varepsilon^r\,\mathrm{d}\varepsilon ,$$

where r is a sufficiently large natural number and $\lambda < 1$.

We show that

$$\lim_{x\to\infty} M(x) = 0 . \tag{6}$$

From (5) we get the equation for $M(x)$

$$M(x) + \int_0^x K(x-\xi)M(\xi)\,\mathrm{d}\xi = C(x) ,$$

where

$$\int_0^\infty |K(x)|\,\mathrm{d}x < \infty ,$$

$$\lim_{x\to\infty} C(x) = 0$$

and the Laplace transform of $K(x)$ is different from (-1) in the right-hand halfplane. Accordingly to the theorem of PALEY and WIENER (4). (6) follows immediately and this completes the proof.

A more detailed paper concerning the same problem will be published in *Studia Mathematica*.

(4) R. E. A. C. PALEY and N. WIENER: *Fourier Transforms in the Complex Domain* (New York, 1934).

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

Some Remarks on the Asymptotic Behaviour of the Cosmic Ray Cascade for Large Depth of the Absorber.

II.

Asymptotic Behaviour of the Probability Distribution Function.

J. ŁOPUSZAŃSKI

Institute of Theoretical Physics of the Polish Academy of Science - Wrocław

(ricevuto il 12 Aprile 1955)

CONTENTS : 1. Introduction. – 2. Formulation of the Problem. – 3. Estimation of $\Gamma(\varepsilon, 0, x)$. – 4. Energetic relations. – 5. Probability Distribution Function. – 6. Conclusions. – Appendix.

1. – Introduction.

In our considerations we confine ourselves for the sake of simplicity to the case of a nucleon cascade in homogeneous nuclear matter. The results given below may be generalized to the case of a nucleon cascade in finite absorber and of an electron-photon cascade.

The development of the nucleon cascade is described by JÁNOSSY's G-equation [1]:

$$\frac{\partial G(\varepsilon, u, x)}{\partial x} + G(\varepsilon, u, x) = \int_0^1 \int_0^1 w(\varepsilon', \varepsilon'')\, \mathrm{d}\varepsilon'\, \mathrm{d}\varepsilon''\, G(\varepsilon/\varepsilon', u, x)\, G(\varepsilon/\varepsilon'', u, x). \tag{1}$$

where

$$G(\varepsilon, u, x) = \sum_{n=0}^{\infty} P(n \mid \varepsilon, x) u^n$$

[1] L. JÁNOSSY: *Proc. Phys. Soc.*, A **63**, 241 (1950).

is the generating function for the probabilities $P(n\,|\,\varepsilon, x)$ to find n nucleons of the cascade with energy $>\varepsilon$ at the depth x, provided that the cascade is due to one nucleon of energy 1, and $w(\varepsilon', \varepsilon'')$ is the cross-section for nucleon-nucleon collision in which the secondary nucleons receive the ε'-th and ε''-th part of the energy of the incident nucleon. We have further

$$w(\varepsilon', \varepsilon'') = 0 \qquad \text{for} \quad \varepsilon' + \varepsilon'' > 1$$

$$w(\varepsilon', \varepsilon'') = 0 \qquad \text{for} \quad \varepsilon' < \varepsilon''$$

$$\int_0^1 \int_0^1 w(\varepsilon', \varepsilon'')\, \mathrm{d}\varepsilon'\, \mathrm{d}\varepsilon'' = 1 \,.$$

The solution of equation (1) in the form of a recurrent formula reads for $\varepsilon > 1/n$ [2]

$$(2) \qquad \Gamma_n(\varepsilon, u, x) = (1-u) S_1(\varepsilon, x) - \int_0^x \mathrm{d}\xi \int_0^1 \left(- \frac{\partial S_1(\eta, x-\xi)}{\partial \eta} \right) \cdot$$

$$\cdot \int_0^1 \int_0^1 w\left(\frac{\varepsilon'}{\eta}, \frac{\varepsilon''}{\eta}\right) \frac{\mathrm{d}\varepsilon'\, \mathrm{d}\varepsilon''}{\eta^2}\, \Gamma_{n-1}\left(\frac{\varepsilon}{\varepsilon'}, u, \xi\right) \Gamma_{n-1}\left(\frac{\varepsilon}{\varepsilon''}, u, \xi\right)$$

where

$$\Gamma_n(\varepsilon, u, x) = 1 - G(\varepsilon, u, x)$$

for $\varepsilon > 1/n$ and $S_1(\varepsilon, x)$ denotes the first moment of the distribution. We outline the derivation of this formal solution (2) in the Appendix.

We denote $\lim_{n\to\infty} \Gamma_n(\varepsilon, u, x)$ by $\Gamma(\varepsilon, u, x)$. The relation $\Gamma(\varepsilon, u, x) = 1 - G(\varepsilon, u, x)$ is valid for all positive ε. Then (2) goes over into an integral equation for $\Gamma(\varepsilon, u, x)$

$$(3) \qquad \Gamma(\varepsilon, u, x) = (1-u) S_1(\varepsilon, x) - \int_0^x \mathrm{d}\xi \int_0^1 \left(- \frac{\partial S_1(\eta, x-\xi)}{\partial \eta} \right) \cdot$$

$$\cdot \int_0^1 \int_0^1 w\left(\frac{\varepsilon'}{\eta}, \frac{\varepsilon''}{\eta}\right) \frac{\mathrm{d}\varepsilon'\, \mathrm{d}\varepsilon''}{\eta^2}\, \Gamma\left(\frac{\varepsilon}{\varepsilon'}, u, \xi\right) \Gamma\left(\frac{\varepsilon}{\varepsilon''}, u, \xi\right) .$$

Equation (3) is equivalent with the integral equation given by JÁNOSSY. It is suitable for our purposes.

It is very tedious to get numerical results from solution (2). We show.

---

[2] J. ŁOPUSZAŃSKI: *Acta Phys. Pol.* **12**, 156 (1953); **14**, 191 (1955).

however, in the next paper that for large depth of the absorber we are able to derive from this solution relatively simple expressions for the distribution function. To this aim we must consider two problems dealt with in this note.

2. – Formulation of the problem.

It follows immediately from (3) that for each treshold energy $\varepsilon > 0$ the cascade must die out. We have namely

$$0 \leqslant \lim_{x\to\infty} \Gamma(\varepsilon, u, x) \leqslant (1-u) \lim_{x\to\infty} S_1(\varepsilon, x) = 0$$

(with respect to the last step see [3]), consequently

$$\lim_{x\to\infty} G(\varepsilon, u, x) = 1$$

and this yields the required result.

It is, however, interesting to know whether the probabilities $P(n \mid \varepsilon, x)$ for $n > 0$ tend to zero equally strong or not. To investigate this problem we introduce the conditional probabilities

$$P^*(n \mid \varepsilon, x) = \frac{P(n \mid \varepsilon, x)}{\sum_{m=1}^{\infty} P(m \mid \varepsilon, x)} = \frac{P(n \mid \varepsilon, x)}{\Gamma(\varepsilon, 0, x)}$$

to find n nucleons of energy $> \varepsilon$ at depth x, provided at least one nucleon of energy $> \varepsilon$ at depth x exists.

We show that

$$\lim_{x\to\infty} \left\{ \frac{\Gamma(\varepsilon', 0, x)}{\Gamma(\varepsilon, 0, x)} \right\} = 0 \tag{4}$$

for $\varepsilon' > \varepsilon > 0$ and

$$\left.\begin{aligned} &\lim_{x\to\infty} P^*(1 \mid \varepsilon, x) = 1 \\ &\lim_{x\to\infty} P^*(n \mid \varepsilon, x) = 0 \quad \text{for } n \geqslant 2 \end{aligned}\right\} \quad \text{for } \varepsilon > 0 \,. \tag{5}$$

The necessary and sufficient condition that (5) should occur is

$$\lim_{x\to\infty} S_1^*(\varepsilon, x) = \lim_{x\to\infty} \left\{ \frac{S_1(\varepsilon, x)}{\Gamma(\varepsilon, 0, x)} \right\} = 1 \,. \tag{6}$$

[3] K. URBANIK: see this issue, pag. 1147.

This becomes evident if we take into account that

$$S_1^*(\varepsilon, x) = 1 + \sum_{n=2}^{\infty} (n-1)P^*(n \mid \varepsilon, x) . \tag{7}$$

The number of terms on the right-hand side of (7) is finite, because for each $\varepsilon > 0$ all $P^*(n \mid \varepsilon, x) = 0$ for $n > [\varepsilon^{-1}]$. Further all this terms are non-negative.

Let us therefore deal with (6).

3. – Estimation of $\Gamma(\varepsilon, 0, x)$.

In paper I it was shown that

$$S_1(\varepsilon, x) = e^{-x} W_1(\varepsilon, x) \tag{8a}$$

where for each $\beta > 0$

$$\lim_{x \to \infty} e^{-\beta x} W_1(\varepsilon, x) = 0 \tag{8b}$$

and

$$\lim_{x \to \infty} W_1(\varepsilon, x) = \infty \tag{8c}$$

for $1 > \varepsilon > 0$. One may extend this result to the function $\Gamma(\varepsilon, u, x)$. For our purposes it is, however, sufficient to show that

$$\Gamma(\varepsilon, 0, x) = e^{-x} V(\varepsilon, x) \tag{9a}$$

where for each $\beta > 0$

$$\lim_{x \to \infty} e^{-\beta x} V(\varepsilon, x) = 0 \tag{9b}$$

and

$$\lim_{x \to \infty} V(\varepsilon, x) = \infty . \tag{9c}$$

To show (9) we notice that

$$\Gamma(\varepsilon, 0, x) \leqslant S_1(\varepsilon, x) \leqslant [1/\varepsilon] \Gamma(\varepsilon, 0, x) . \tag{10}$$

From (8), (9) and (10) we have

$$0 \leqslant \lim_{x \to \infty} e^{-\beta x} V(\varepsilon, x) \leqslant \lim_{x \to \infty} e^{-\beta x} W_1(\varepsilon, x) = 0 \tag{11}$$

and further

$$[1/\varepsilon] \lim_{x \to \infty} V(\varepsilon, x) \geqslant \lim_{x \to \infty} W_1(\varepsilon, x) = \infty \tag{12}$$

(11) and (12) complete the proof of (9).

4. – Energetic relations.

According to (1) $\Gamma(\varepsilon, 0, x)$ satisfies the following equation.

$$\frac{\partial \Gamma(\varepsilon, 0, x)}{\partial x} + \Gamma(\varepsilon, 0, x) = \int_0^1 w(\varepsilon')\,\mathrm{d}\varepsilon'\,\Gamma(\varepsilon/\varepsilon', 0, x) - \int_0^1\int_0^1 \Gamma(\varepsilon/\varepsilon', 0, x)\,\Gamma(\varepsilon/\varepsilon'', 0, x)\,w(\varepsilon', \varepsilon'')\,\mathrm{d}\varepsilon'\,\mathrm{d}\varepsilon'', \tag{13}$$

with

$$w(\varepsilon') = \int_0^1 \{w(\varepsilon', \varepsilon'') + w(\varepsilon'', \varepsilon')\}\,\mathrm{d}\varepsilon'' .$$

Using the notation of (9*a*) the equation (13) takes the form

$$\frac{\partial V(\varepsilon, x)}{\partial x} = \int_0^1 w(\varepsilon')\,\mathrm{d}\varepsilon'\,V(\varepsilon/\varepsilon', x) - e^{-x}\int_0^1\int_0^1 V(\varepsilon/\varepsilon', x)\,V(\varepsilon/\varepsilon'', x)\,w(\varepsilon', \varepsilon'')\,\mathrm{d}\varepsilon'\,\mathrm{d}\varepsilon''. \tag{14}$$

Let us divide both sides of (14) by $V(\varepsilon, x)$.

Since

$$e^{-x}\int_0^1\int_0^1 \frac{V(\varepsilon/\varepsilon', x)\,V(\varepsilon/\varepsilon'', x)}{V(\varepsilon, x)}\,w(\varepsilon', \varepsilon'')\,\mathrm{d}\varepsilon'\,\mathrm{d}\varepsilon'' = \int_0^1 \mathrm{d}\varepsilon'\,V(\varepsilon/\varepsilon', x)\,e^{-x}\int_0^1 \mathrm{d}\varepsilon''\,w(\varepsilon', \varepsilon''),$$

the last term on the right-hand side of (14) tends to zero if x tends to infinity.

We make now the assumption that the limit of

$$\frac{1}{V(\varepsilon, x)}\,\frac{\partial V(\varepsilon, x)}{\partial x}$$

exists if x tends to infinity. We justify this assumption in the following manner. There are only two possibilities, namely either the limit exists or not. We consider the first case, because this case seems to us physically more probable. We denote this limit by

$$A = \lim_{x\to\infty}\left\{\frac{1}{V(\varepsilon, x)}\,\frac{\partial V(\varepsilon, x)}{\partial x}\right\}.$$

In respect to (9c) A can not be negative. On the other hand from (9b) it follows that A can not be positive too. Consequently it must be equal to 0.

In this way our equation (14) reduces to

$$\lim_{x\to\infty}\left\{\int_0^1 w(\varepsilon')\,\mathrm{d}\varepsilon'\,\frac{V(\varepsilon/\varepsilon', x)}{V(\varepsilon, x)}\right\} = 0\,. \tag{15}$$

The sufficient but not necessary condition that the limit and integration operators commute, is the uniform convergence of

$$\frac{V(\varepsilon/\varepsilon', x)}{V(\varepsilon, x)}$$

when x tends to infinity. Our expression does not fulfill this condition. Nevertheless we are able to show that the interchange of operators is permissible. We get from (15)

$$\int_0^1 w(\varepsilon')\,\mathrm{d}\varepsilon' \lim_{x\to\infty}\left\{\frac{V(\varepsilon/\varepsilon', x)}{V(\varepsilon, x)}\right\} = 0$$

and finally, making use of that $\Gamma(\varepsilon, 0, x)$ is a nonincreasing function of ε,

$$\lim_{x\to\infty}\left\{\frac{\Gamma(\varepsilon', 0, x)}{\Gamma(\varepsilon, 0, x)}\right\} = 0 \tag{4}$$

for $\varepsilon' > \varepsilon > 0$. We have the result stated.

5. – Probability distribution function.

We show now that (4) yields (6).

To this aim we must make use of our integral equation (3). We put $u = 0$ and divide it by $\Gamma(\varepsilon, 0, x)$. We get

$$1 = S_1^*(\varepsilon, x) - \int_0^x \mathrm{d}\xi \int_0^1 \mathrm{d}\eta\left(-\frac{\partial S_1(\eta, x-\xi)}{\partial\eta}\right)\cdot \tag{16}$$

$$\cdot \int_0^1\int_0^1 w\left(\frac{\varepsilon'}{\eta}, \frac{\varepsilon''}{\eta}\right)\frac{\mathrm{d}\varepsilon'\,\mathrm{d}\varepsilon''}{\eta^2}\,\frac{\Gamma(\varepsilon/\varepsilon', 0, \xi)\,\Gamma(\varepsilon/\varepsilon'', 0, \xi)}{\Gamma(\varepsilon, 0, x)}\,.$$

Let us consider the second term on the right-hand side of (16). We show that this expression tends to zero if x tends to infinity and in this way complete the proof.

We write this term in the form

$$D(\eta, \varepsilon', \varepsilon'')\int_0^x \mathrm{d}\xi \frac{-\partial S_1(\eta, x-\xi)/\partial\eta}{\Gamma(\varepsilon, 0, x)} \cdot \Gamma\left(\frac{\varepsilon}{\varepsilon'}, 0, \xi\right) \Gamma\left(\frac{\varepsilon}{\varepsilon''}, 0, \xi\right) \tag{17}$$

where $D(\eta, \varepsilon', \varepsilon'')$ denotes the operator

$$\int_0^1 \mathrm{d}\eta \int_0^1\int_0^1 \frac{\mathrm{d}\varepsilon'\,\mathrm{d}\varepsilon''}{\eta^2}\, w\left(\frac{\varepsilon'}{\eta}, \frac{\varepsilon''}{\eta}\right).$$

Using the notation of (8) and (9) the term (17) goes over in

$$D(\eta, \varepsilon', \varepsilon'')\int_0^x \mathrm{d}\xi \frac{-\partial W_1(\eta, x-\xi)/\partial\eta}{V(\varepsilon, x)} \cdot V\left(\frac{\varepsilon}{\varepsilon'}, \xi\right) V\left(\frac{\varepsilon}{\varepsilon''}, \xi\right) e^{-\xi}. \tag{18}$$

It follows from the differential equation satisfied by $-\partial W_1(\eta, x)/\partial\eta$ that $-\partial W_1(\eta, x)/\partial\eta$ is a nondecreasing function of increasing x. This equation reads

$$\frac{\partial}{\partial x}\left(-\frac{\partial W_1(\eta, x)}{\partial\eta}\right) = \int_0^1 \frac{w(\varepsilon')}{\varepsilon'}\,\mathrm{d}\varepsilon'\left(-\frac{\partial W_1(\eta/\varepsilon', x)}{\partial(\eta/\varepsilon')}\right).$$

Consequently

$$(18) \leqslant D(\eta, \varepsilon', \varepsilon'')\frac{-\partial W_1(\eta, x)/\partial\eta}{V(\varepsilon, x)}\int_0^x V\left(\frac{\varepsilon}{\varepsilon'}, \xi\right) V\left(\frac{\varepsilon}{\varepsilon''}, \xi\right) e^{-\xi}\,\mathrm{d}\xi. \tag{19}$$

Let us estimate the integral

$$0 \leqslant \int_0^\infty V\left(\frac{\varepsilon}{\varepsilon'}, \xi\right) e^{-\xi/3} V\left(\frac{\varepsilon}{\varepsilon''}, \xi\right) e^{-\xi/3} e^{\xi/3}\,\mathrm{d}\xi \cdot$$

$$\leqslant 3 \max_{0\leqslant\xi\leqslant\infty}\left(V\left(\frac{\varepsilon}{\varepsilon'}, \xi\right) e^{-\xi/3}\right) \max_{0\leqslant\xi\leqslant\infty}\left(V\left(\frac{\varepsilon}{\varepsilon''}, \xi\right) e^{-\xi/3}\right)$$

and the last expression is bounded because of (9*b*). Due to this circumstance the right-hand side of (19) vanishes for x tending to infinity, if

$$\lim_{x\to\infty}\left\{\frac{-\partial W_1(\eta, x)/\partial\eta}{V(\varepsilon, x)}\right\} = 0.$$

We take into account that

$$\lim_{x\to\infty}\left\{\frac{S_1(\eta, x)}{\Gamma(\varepsilon, 0, x)}\right\} = 0$$

for $\eta > \varepsilon$ occurs because of (4). In order to show that the relations

$$0 = -\frac{\partial}{\partial \eta}\left|\lim_{x\to\infty}\frac{S_1(\eta, x)}{\Gamma(\varepsilon, 0, x)}\right| = \lim_{x\to\infty}\left\{-\frac{\partial S_1(\eta, x)/\partial\eta}{\Gamma(\varepsilon, 0, x)}\right\} \qquad \text{for } \eta > \varepsilon \tag{20}$$

hold, it was sufficient to show that

$$\frac{S_1(\eta, x)}{\Gamma(\varepsilon, 0, x)},$$

converges for x tending to infinity, has continuous derivative in respect to η and that this derivative converges uniformly for x tending to infinity. The dangerous places of expression (20) lie in the neighbourhood of $\eta_k = 1,\ 1/2,\ 1/3, \ldots$, e.g.

$$\frac{\partial W_1\,(\eta = 1,\ x)}{\partial \eta}$$

shows Dirac-delta singularity for all x. But we are able to overcome this difficulty. We outline only the proof because this matter is rather interesting for a mathematician than for a physicist. We divide

$$\frac{W_1(\eta, x)}{\Gamma(\varepsilon, 0, x)}$$

into two parts: the continuous one

$$\frac{W_1^{(c)}(\eta, x)}{\Gamma(\varepsilon, 0, x)}$$

and a step function

$$\frac{W_1^{(s)}(\eta, x)}{\Gamma(\varepsilon, 0, x)} = \frac{1}{\Gamma}\sum_{k=1}^{[1/\varepsilon]} a_k(x)\,H(\eta - \eta_k)$$

where $H(z)$ denotes the Heaviside's function. We consider these two functions separately. We put

$$\frac{\partial W_1}{\partial \eta} = \frac{\partial W_1^{(c)}}{\partial \eta} + \frac{\partial W_1^{(s)}}{\partial \eta}$$

into the right-hand side of (19). What belongs the singular part we are able to perform the integration in respect to η and show that this part vanishes for x tending to infinity. We notice that the number of singularities is finite for each $\varepsilon > 0$, namely $< [1/\varepsilon]$. The continuous function

$$\frac{\partial W_1^{(c)}(\eta, x)/\partial\eta}{\Gamma(\varepsilon, 0, x)}$$

leads to no difficulties: (20) holds for it.

For $\eta < \varepsilon' + \varepsilon''$ as well for $\varepsilon' < \varepsilon$ and $\varepsilon'' < \varepsilon$ expression (17) vanishes, consequently we employ in (18) as well as in the right-hand side of (19) the value of $\eta > 2\varepsilon$ only.

Let x in formula (19) tend to infinity. Then

$$D(\eta, \varepsilon', \varepsilon'') \lim_{x\to\infty} \left\{ - \frac{\partial W_1^{(e)}(\eta, x)/\partial\eta}{V(\varepsilon, x)} \right\} \cdot \int_0^\infty V\left(\frac{\varepsilon}{\varepsilon'}, \xi\right) V\left(\frac{\varepsilon}{\varepsilon''}, \xi\right) e^{-\xi} \, \mathrm{d}\xi = 0$$

and we have the result stated.

In this way we have completed the proof of (5) and we can finally write for large x

$$\begin{cases} P(0 \mid \varepsilon, x) \sim 1 \\ P(1 \mid \varepsilon, x) \sim S_1(\varepsilon, x) \\ P(n \mid \varepsilon, x) = o(P(1 \mid \varepsilon, x)) \qquad \text{for } n > 1. \end{cases} \tag{21}$$

6. – Conclusions.

We have proved (4) and (5). $\Gamma(\varepsilon', 0, x)/\Gamma(\varepsilon, 0, x)$ denotes the conditional probability that we find at least one nucleon of energy $> \varepsilon'$, provided that there is at least one nucleon of energy $> \varepsilon$ where $\varepsilon' > \varepsilon$ at depth x.

Let us now consider the physical meaning of these two theorems.

The relation (4) states: for very large depth of the absorber if we find at least one nucleon of energy $> \varepsilon$ we may be sure that the energy of the nucleons lies in the infinitesimal energy range about ε i.e. $(\varepsilon, \varepsilon + \mathrm{d}\varepsilon)$.

The relation (5) or (21) states: for very large depth of the absorber if we find at least one nucleon of energy $> \varepsilon$ we may be sure that there is only one nucleon.

Both relations taken together state: if we find for very large depth of the absorber at least one nucleon of energy $> \varepsilon$, we may be sure that there is only one nucleon his energy lying in the infinitesimal energy range $(\varepsilon, \varepsilon + d\varepsilon)$.

Formulae (5) or (21) give asymptotic solutions for the distribution function. This result may be interesting from the point of view of the theoretical physicist. The experimental physicist, however, requires among other numerical data for large but finite depth of the absorber. Our solution realizes the first step in an approximation procedure for large but finite x. It seems very probable that the next step in the approximation will be given by the second factorial moment, the next following by the third and so on. The second part of this paper is devoted to this problem.

* * *

The author wishes to express his sincere thanks to Professor J. Rzewuski for continued encouragement and helpful discussions and to Mgr. K. Urbanik and H. Stackowiak for helpful discussions.

Appendix

The G-equation reads

$$\frac{\partial G(\varepsilon, u, x)}{\partial x} + G(\varepsilon, u, x) = \int_0^1 \int_0^1 w(\varepsilon', \varepsilon'')\, \mathrm{d}\varepsilon'\, \mathrm{d}\varepsilon''\, G\left(\frac{\varepsilon}{\varepsilon'}, u, x\right) G\left(\frac{\varepsilon}{\varepsilon''}, u, x\right). \tag{1}$$

The boundary conditions are

$$G(\varepsilon, u, x = 0) = \begin{cases} u & \text{for } \varepsilon \leqslant 1, \\ 1 & \text{for } \varepsilon > 1. \end{cases} \tag{1.A}$$

For $\varepsilon > 1$, (1) and (1.A) will be of course satisfied by

$$G(\varepsilon, u, x) = 1.$$

For $\varepsilon > \frac{1}{2}$ we introduce an auxiliary function

$$\Gamma_2(\varepsilon, u, x) = 1 - G(\varepsilon, u, x).$$

We express G in (1) in terms of Γ_2 and find

$$\frac{\partial \Gamma_2(\varepsilon, u, x)}{\partial x} + \Gamma_2(\varepsilon, u, x) = \int_0^1 \mathrm{d}\varepsilon'\, w(\varepsilon')\, \Gamma_2\left(\frac{\varepsilon}{\varepsilon'}, u, x\right) - \int_0^1 \int_0^1 \mathrm{d}\varepsilon'\, \mathrm{d}\varepsilon''\, w(\varepsilon', \varepsilon'')\, \Gamma_2\left(\frac{\varepsilon}{\varepsilon'}, u, x\right) \Gamma_2\left(\frac{\varepsilon}{\varepsilon''}, u, x\right). \tag{2.A}$$

If we take into account that

$$w(\varepsilon', \varepsilon'') = 0 \quad \text{for} \quad \varepsilon' + \varepsilon'' > 1$$

and

$$\Gamma\left(\frac{\varepsilon}{\varepsilon'}\right) \Gamma\left(\frac{\varepsilon}{\varepsilon''}\right) = 0 \quad \text{for} \quad \varepsilon' < \varepsilon \quad \text{or} \quad \varepsilon'' < \varepsilon$$

then it becomes obvious that the last term vanishes. (2.A) reduces to

$$\frac{\partial \Gamma_2(\varepsilon, u, x)}{\partial x} + \Gamma_2(\varepsilon, u, x) = \int_0^1 w(\varepsilon')\mathrm{d}\varepsilon' \Gamma_2\left(\frac{\varepsilon}{\varepsilon'}, u, x\right). \tag{3.A}$$

But (3.A) has the form of the equation for the first moment [4] only the boundary conditions are different. To solve (3.A) we apply the well known method of Mellin transform and find

$$\Gamma_2(\varepsilon, u, x) = (1-u)\, S_1(\varepsilon, x). \tag{4.A}$$

For $\varepsilon > \frac{1}{3}$ we introduce

$$\Gamma_3(\varepsilon, u, x) = 1 - G(\varepsilon, u, x)$$

and apply the same procedure as in the case of function $\Gamma_2(\varepsilon, u, x)$. We obtain the equation

$$\frac{\partial \Gamma_3(\varepsilon, u, x)}{\partial x} + \Gamma_2(\varepsilon, u, x) - \int_0^1 \Gamma_3\left(\frac{\varepsilon}{\varepsilon'}, u, x\right) w(\varepsilon')\mathrm{d}\varepsilon' = \tag{5.A}$$

$$= -\int_0^1\int_0^1 \Gamma_2\left(\frac{\varepsilon}{\varepsilon'}, u, x\right)\Gamma_2\left(\frac{\varepsilon}{\varepsilon''}, u, x\right) w(\varepsilon', \varepsilon'')\,\mathrm{d}\varepsilon'\,\mathrm{d}\varepsilon''.$$

Generally, for $\varepsilon > 1/n$, we get for

$$\Gamma_n(\varepsilon, u, x) = 1 - G(\varepsilon, u, x)$$

an equation analogous to (5.A), namely

$$\frac{\partial \Gamma_n(\varepsilon, u, x)}{\partial x} + \Gamma_n(\varepsilon, u, x) - \int_0^1 \Gamma_n\left(\frac{\varepsilon}{\varepsilon'}, u, x\right) w(\varepsilon')\,\mathrm{d}\varepsilon' = \tag{6.A}$$

$$= -\int_0^1\int_0^1 \Gamma_{n-1}\left(\frac{\varepsilon}{\varepsilon'}, u, x\right)\Gamma_{n-1}\left(\frac{\varepsilon}{\varepsilon''}, u, x\right) w(\varepsilon', \varepsilon'')\,\mathrm{d}\varepsilon'\,\mathrm{d}\varepsilon''.$$

We are able to solve (6.A) by means of the Mellin transform [4], provided we hold $\Gamma_{n-1}(\varepsilon, u, x)$ to be a known function. The solution of (6.A) is simultaneously the formal solution of (1). It has the form given by (2).

This Appendix is a summary of a paper submitted to the *Acta Physica Polonica* [2].

[4] H. Messel: *Proc. Roy. Irish Acad.*, A **54**, 125 (1951).

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

Some Remarks on the Asymptotic Behaviour of the Cosmic Ray Cascade for Large Depth of the Absorber.

III.

Evaluation of the Distribution Function.

J. ŁOPUSZAŃSKI

Institute of Theoretical Physics of the Polish Academy of Science - Wrocław

(ricevuto il 12 Aprile 1955)

CONTENTS : 1. Introduction. – 2. The Behaviour of the Factorial Moments. – 3. The Second and Higher Approximations to the Probability Distribution Function. – 4. Conclusions.

1. – Introduction.

In a previous paper (ŁOPUSZAŃSKI [1]), hereafter referred to as paper II) we have obtained asymptotic formulae for the probability distribution function for the nucleon cascade in homogeneous nuclear matter

$$(1)\qquad \begin{cases} P(0\,|\,\varepsilon, x) \sim 1 \\ P(1\,|\,\varepsilon, x) \sim S_1(\varepsilon, x) \\ P(n\,|\,\varepsilon, x) = o\big(P(1\,|\,\varepsilon, n)\big) \qquad \text{for } n > 1 \end{cases}$$

(for notation see paper II). One may consider (1) as the first approximation for the distribution function for large but finite x.

In paper II we have also shown that

$$(2)\qquad \lim_{x\to\infty} \left\{ \frac{\Gamma(\varepsilon', 0, x)}{\Gamma(\varepsilon, 0, x)} \right\} \qquad 0 \qquad \text{for } \varepsilon' < \varepsilon .$$

[1] J. ŁOPUSZAŃSKI: see this issue, pag. 1150.

To investigate the next approximations we consider the probabilities of finding at least two, three and s.o. nucleons of energy $> \varepsilon$ at the depth x.

It is well known that

$$G(\varepsilon, u, x) = 1 + \sum_{k=1}^{\infty} (-1)^k \frac{1}{k!} S_k(\varepsilon, x)\,(1-u)^k$$

where $S_k(\varepsilon, x) = \sum_{n=0}^{\infty} n(n-1)\ldots(n-k+1)P(n \mid \varepsilon, x)$ denotes the k-th factorial moment. Hence

$$\Gamma(\varepsilon, u, x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k!} S_k(\varepsilon, x)\,(1-u)^k . \tag{3}$$

One may easily prove the following relation

$$\frac{\Gamma(\varepsilon, u, x)}{1-u} = [1 - P(0 \mid \varepsilon, x)] + [1 - P(0 \mid \varepsilon, x) - P(1 \mid \varepsilon, x)]u + \\ + [1 - P(0 \mid \varepsilon, x) - P(1 \mid \varepsilon, x) - P(2 \mid \varepsilon, x)]u^2 + \ldots .$$

This formula enables us to write down at once the expressions for the probabilities to find at least k ($k = 1, 2, \ldots$) nucleons of energy $> \varepsilon$ at depth x:

$$\sum_{l=k}^{\infty} P(l \mid \varepsilon, x) = \frac{1}{(k-1)!} \left[\frac{\partial^{k-1}}{\partial u^{k-1}} \frac{\Gamma(\varepsilon, u, x)}{1-u} \right]_{u=0} = \sum_{l=k}^{\infty} (-1)^{l-k} \frac{S_l(\varepsilon, x)}{l!} \binom{l}{k-1} . \tag{4}$$

2. – The Behaviour of the Factorial Moments.

In paper II we have presented a new integral equation equivalent to the integral G-equation of JÁNOSSY [2]

$$\Gamma(\varepsilon, u, x) = (1-u)\,S_1(\varepsilon, x) - D(\eta, \varepsilon', \varepsilon'') \cdot \\ \cdot \int_0^x \mathrm{d}\xi \left(- \frac{\partial S_1(\eta, x-\xi)}{\partial \eta} \right) \cdot \Gamma\left(\frac{\varepsilon}{\varepsilon'}, u, \xi\right) \Gamma\left(\frac{\varepsilon}{\varepsilon''}, u, \xi\right) . \tag{5}$$

If we substitute for the Γ's on the right-hand side of (5) the expressions for Γ, given by (5), and repeat this procedure infinitely many times, we get a formal expansion of Γ in powers of $(1-u)$. The coefficient of $(1-u)^2$ reads e.g.

$$\frac{1}{2} S_2(\varepsilon, x) = D(\eta, \varepsilon', \varepsilon'') \int_0^x \mathrm{d}\xi \left(- \frac{\partial S_1(\eta, x-\xi)}{\partial \eta} \right) \cdot S_1\left(\frac{\varepsilon}{\varepsilon'}, \xi\right) S_1\left(\frac{\varepsilon}{\varepsilon''}, \xi\right) \tag{6}$$

[2] L. JÁNOSSY: *Proc. Phys. Soc.*, A **63**, 241 (1950).

and generally, the coefficient of $(1-u)^n$

$$\frac{1}{n!}\,S_n(\varepsilon, x) = D(\eta, \varepsilon', \varepsilon'')\int_0^x d\xi\left(-\frac{\partial S_1(\eta, x-\xi)}{\partial \eta}\right)\cdot \sum_{k=1}^{n-1}\frac{1}{k!}\,S_k\left(\frac{\varepsilon}{\varepsilon'}, \xi\right)\frac{1}{(n-k)!}\,S_{n-k}\left(\frac{\varepsilon}{\varepsilon''}, \xi\right). \tag{7}$$

We show now the important relation

$$\lim_{x\to\infty}\left\{\frac{S_n(\varepsilon, x)}{S_m(\varepsilon, x)}\right\} = 0 \qquad \text{for } n > m \quad \text{and} \quad \varepsilon < 0. \tag{8}$$

To this aim let us prove the auxiliary lemma, that

$$\lim_{x\to\infty}\left\{-\frac{\partial S_1(\omega, x)/\partial\omega}{S_2(\omega', x)}\right\} = 0 \qquad \text{for } \omega > 2\omega'. \tag{9}$$

To prove formula (9) we take advantage of the fact that we may find such a sufficiently large, finite $X(\omega')$ that for $x > X$ the function $-\partial S_1(\eta, x)/\partial\eta$ is a non-increasing function for all $\eta \geqslant 2\omega' > 0$. Consequently we have

$$\frac{1}{2}\,S_2(\omega', x) - D(\eta, \varepsilon', \varepsilon'')\left\{\int_0^{x-X(\omega')} d\xi\left(-\frac{\partial S_1(\eta, x-\xi)}{\partial\eta}\right)\cdot S_1\left(\frac{\omega'}{\varepsilon'}, \xi\right) S_1\left(\frac{\omega'}{\varepsilon''}, \xi\right) + \int_{x-X(\omega')}^{x} d\xi\left(-\frac{\partial S_1(\eta, x-\xi)}{\partial\eta}\right)\cdot S_1\left(\frac{\omega'}{\varepsilon'}, \xi\right) S_1\left(\frac{\omega'}{\varepsilon''}, \xi\right)\right\} > D(\eta, \varepsilon', \varepsilon'')\left(-\frac{\partial S_1(\eta, x)}{\partial\eta}\right)\cdot \int_0^{x-X(\omega')} d\xi\, S_1\left(\frac{\omega'}{\varepsilon'}, \xi\right) S_1\left(\frac{\omega'}{\varepsilon''}, \xi\right).$$

Hence

$$\lim_{x\to\infty}\left\{\frac{\frac{1}{2}S_2(\omega', x)}{-\partial S_1(\omega, x)/\partial\omega}\right\} > D(\eta, \varepsilon', \varepsilon'')\lim_{x\to\infty}\left\{\frac{\partial S_1(\eta, x)/\partial x}{\partial S_1(\omega, x)/\partial\omega}\right\}\cdot \int_0^{\infty} d\xi\, S_1\left(\frac{\omega'}{\varepsilon'}, \xi\right) S_1\left(\frac{\omega'}{\varepsilon''}, \xi\right). \tag{10}$$

It was shown in paper II that (2) yields

$$\lim_{x\to\infty}\left\{-\frac{\partial S_1(\omega, x)/\partial\omega}{F(\eta, 0, x)}\right\} = 0 \qquad \text{for } \omega > \eta > 0.$$

Hence

$$\lim_{x\to\infty}\left\{\frac{\partial S_1(\omega, x)/\partial\eta}{\Gamma(\eta, 0, x)}\right\}$$

is different from zero and singular. Consequently we have

$$\lim_{x\to\infty}\left\{\frac{-\partial S_1(\omega, x)/\partial\omega}{\Gamma(\eta, 0, x)}\right\}\lim_{x\to\infty}\left\{\frac{\Gamma(\eta, 0, x)}{-\partial S_1(\eta, x)/\partial\eta}\right\} = \lim_{x\to\infty}\left\{\frac{\partial S_1(\omega, x)/\partial\omega}{\partial S_1(\eta, x)/\partial\eta}\right\} = 0 \tag{11}$$

for $\omega > 2\eta$ [3].

The integral on the right-hand side of (10) is positive because the subintegral functions are positive and continuous. Due to the properties of $D(\eta, \varepsilon', \varepsilon'')$ (cf. paper II) η in (10) takes on values from the interval $(2\omega', 1)$. If in this interval a small but finite subinterval exists for which $\eta < \omega$ then by virtue of (11), the right-hand side of (10) tends to infinity. In this case we have $2\omega' < \omega$ and the result follows.

The formula (9) can be generalized to

$$\lim_{x\to\infty}\left\{\frac{-\partial S_1(\omega, x)/\partial\omega}{S_n(\omega', x)}\right\} = 0 \qquad \text{for} \quad \omega > n\omega' \tag{12}$$

The proof is similar as in the case of (9).

Starting from (9) we can deduce the following relations

$$\lim_{x\to\infty}\left\{\frac{S_n(\varepsilon, x)}{S_2(\varepsilon, x)}\right\} = 0 \qquad \text{for} \quad n > 2. \tag{13}$$

Let us prove this relation. It is well known that $S_n(\varepsilon, x) = 0$ for $\varepsilon > 1/n$. This becomes at once evident from (6) and (7) if we take into account that $S_1(\varepsilon, x) = 0$ for $\varepsilon > 1$. We have for the integration intervals of the variables appearing in (6) and (7) the relations

$$1 \geqslant \eta \geqslant \varepsilon' + \varepsilon'' \geqslant 2\varepsilon \quad \text{and} \quad 1 \geqslant \eta \geqslant \varepsilon' + \varepsilon'' \geqslant n\varepsilon \quad \text{resp.}$$

It was shown (URBANIK [4]), hereafter referred to as paper I) that for $0 < \varepsilon < 1$

$$S_k(\varepsilon, x) = e^{-x} W_k(\varepsilon, x) \qquad k = 1, 2, \ldots \left[\frac{1}{\varepsilon}\right] \tag{14a}$$

[3] Our way to (11) was mathematically not rigorous. If we want to proceed more rigorously we must first prove that the operators $\partial/\partial\omega$ and $\lim_{x\to\infty}$ commute. To show it we proceed in a similar manner as was indicated in paper II in analogous cases.

[4] K. URBANIK: see this issue, pag. 1147.

where

$$\lim_{x\to\infty}\{e^{-\beta x}W_k(\varepsilon, x)\} = 0 \qquad \text{for } \beta > 0 \tag{14b}$$

and

$$\lim_{x\to\infty} W_k(\varepsilon, x) = \infty . \tag{14c}$$

Using (14) we can write

$$0 \leqslant \frac{(1/n!)\,S_n(\varepsilon, x)}{S_2(\varepsilon, x)} \leqslant D(\eta, \varepsilon', \varepsilon'')\left(\frac{-\partial S_1(\eta, x)/\partial x}{S_2(\varepsilon, x)}\right)\cdot \tag{15}$$

$$\cdot\int_0^x \mathrm{d}\xi \sum_{k=1}^{n-1}\frac{1}{k!}\,W_k\left(\frac{\varepsilon}{\varepsilon'}, \xi\right)\frac{1}{(n-k)!}\,W_{n-k}\left(\frac{\varepsilon}{\varepsilon''}, \xi\right)e^{-\xi}.$$

The integral on the right-hand side of (15) is bounded, because of (14), and positive for all x. The right-hand side of (15) vanishes for $x \to \infty$ if

$$\lim_{x\to\infty}\left\{\frac{\partial S_1(\eta, x)/\partial x}{S_2(\varepsilon, x)}\right\} = 0 \tag{16}$$

and this takes place for $\eta > 2\varepsilon$, according to (9). The relation (16) must hold for all η from the interval $(n\varepsilon, 1)$, consequently it is necessary that

$$n\varepsilon \geqslant 2\varepsilon$$

and the result (13) follows.

We may easily generalize the result (13) to (8):

$$\lim_{x\to\infty}\left\{\frac{S_n(\varepsilon, x)}{S_m(\varepsilon, x)}\right\} = 0 \qquad \text{for } n > m \text{ and } \varepsilon > 0.$$

That is our main result. It may be justified in a similar manner as (13), making use of (12).

3. – The Second and Higher Approximations of the Probability Distribution Function.

Formula (8) involves in fact the solution of the problem of evaluation of higher approximations. Indeed, we have, according to (4),

$$\sum_{l=2}^{\infty} P(l \mid \varepsilon, x) = \frac{1}{2}\,S_2(\varepsilon, x) - \frac{2}{3!}\,S_3(\varepsilon, x) + \frac{3}{4!}\,S_4(\varepsilon, x) \quad \ldots . \tag{17}$$

Let us divide both sides of (17) by $\frac{1}{2}S_2(\varepsilon, x)$ and let x tend to infinity. Taking (8) into account, we get

$$\lim_{x\to\infty}\left\{\frac{\sum_{l=2}^{\infty} P(l\,|\,\varepsilon, x)}{\frac{1}{2}S_2(\varepsilon, x)}\right\} = 1. \tag{18}$$

We notice that the number of terms on the right-hand side of (17) does not exceed $[1/\varepsilon]$. It follows immediately from (18) that

$$\begin{cases} P(2\,|\,\varepsilon, x) \sim \frac{1}{2}S_2(\varepsilon, x) \\ P(n\,|\,\varepsilon, x) = o(P(2\,|\,\varepsilon, x)) & \text{for } n \geqslant 3. \end{cases} \tag{19}$$

In the same manner we obtain generally from (4)

$$P(n\,|\,\varepsilon, x) \sim \frac{1}{n!}S_n(\varepsilon, x) \tag{20}$$

$$P(n\,|\,\varepsilon, x) = o(P(m\,|\,\varepsilon, x)) \tag{21}$$

for $n > m$, $n, m = 1, 2, \ldots$ and $\varepsilon > 0$.

5. – Conclusions.

Formulae (20) together with (8) constitute the main result of our note.

It becomes obvious from (3) and (20) that for large x the successive terms of the expansion of $\Gamma(\varepsilon, n, x)$ represent the successive approximation steps.

Equation (3) yields

$$P(n\,|\,\varepsilon, x) = \sum_{k=n}^{\infty}(-1)^{k-n}\binom{k}{n}\frac{1}{k!}S_k(\varepsilon, x). \tag{22}$$

To obtain the expansion of $\Gamma(\varepsilon, n, x)$ in terms of $S_n(\varepsilon, x)$ it is convenient to use either the formula (5) or (7). The series (22) are the s.c. Leibniz series. Taking (8) into account the ratio of the $(n+1)$-th to n-th term tends to zero if $x \to \infty$, consequently a fairly good convergence for large but finite x may be expected. The fact that (22) are Leibniz series admits at once an estimation of the series. We have e.g. for the function $\Gamma(\varepsilon, 0, x)$, which is of great importance in practical work, from (3) successively

$$S_1 \geqslant \Gamma(\varepsilon, 0, x) \geqslant S_1 - \frac{1}{2}S_2, \tag{23a}$$

$$S_1 - \frac{1}{2}S_2 + \frac{1}{6}S_3 \geqslant \Gamma(\varepsilon, 0, x) \geqslant S_1 - \frac{1}{2}S_2 + \frac{1}{6}S_3 - \frac{1}{24}S_4. \tag{23b}$$

The first and second moments were exhaustively investigated by many authors (for nucleon cascade in homogeneous nuclear matter see e.g. ([5,6])). Also higher moments were already considered in some papers (see e.g. ([7])). The results of these investigations enable us to estimate the distribution function according to (23). One may ask for which x the approximation (23*a*) and for which x (23*b*) or a better approximation is fittest. We are not able to answer this question but at any rate (23*a*) certainly does not fit for this x-interval where $S_1(\varepsilon, x) > 1$. The same is also true *mutatis mutandis* for (23*b*).

Further, the question may arise, why the customary approximation method of Picard was not applied to the integral G-equation. This equation admits yet the application of this method, in accordance with the « fix-point-theorem ». Although this method seems to be a straightforward one, it is, however, not suitable for large x because the successive approximation steps consist of sums of graph representations of all possible cascade shapes (trees). It is, however, quite possible that, if we chose as the starting point for the Picard's approximation method the equation (5) instead of G-equation and as the first approximation $-S_1$, we get our results.

We hope that our method will be useful for practical purposes, especially for numerical computations of the distribution function for large depth of the absorber.

The considerations given above can be generalized to the case of the nucleon cascade in finite absorber as well as to the case of the electron-photon cascade.

* * *

The author wishes to express his sincere thanks to Professor J. Rzewuski for continued encouragment and helpful discussions and to Mgr. K. Urbanik for helpful discussions.

([5]) H. Messel: *Proc. Roy. Irish Acad.*, A **54**, 125 (1951).
([6]) H. Messel: *Proc. Phys. Soc.*, A **64**, 726 (1951).
([7]) H. Messel: *Proc. Phys. Soc.*, A **65**, 465 (1952).

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 4, 1955
2° Semestre

Thermodynamical Theory of Galvanomagnetic and Thermomagnetic Phenomena (*).

R. FIESCHI

Gruppo di Fisica dei Solidi, Istituto Nazionale di Fisica Nucleare - Sezione di Milano

(ricevuto il 26 Maggio 1955)

In a previous paper (1), hereafter denoted as (I), the equations describing the galvanomagnetic and thermomagnetic phenomena in anisotropic crystals have been derived from more general macroscopic equations for transfer phenomena in the presence of a magnetic field. The equations which have been obtained (cfr. I, eqs. (II.138, 139) or (II.142, 143)), are slightly different from those used by other authors (2,3), as the gradient of the difference between the electrochemical potentials of electrons and ions, $\tilde{\mu}_e - \tilde{\mu}_i$, appears, instead of the gradient of $\tilde{\mu}_e$ alone (4). It is however easy to show that the two sets of equations are equivalent.

The equations of I are written with respect to the center of mass of the system, i.e. the entropy flow $\boldsymbol{J}_s$ and the electronic flow $\boldsymbol{J}_e$ are defined with respect to the center of mass. If we remember that we are dealing with systems consisting of a rigid ion lattice and of electrons, and that the electronic mass is negligible in comparison with the ionic mass, we see that no physical difference is introduced in writing equations in which the flows are defined with respect to the rigid lattice. The new equations obtained with this change

(*) Appendix to a previous paper. (See reference (1) below).

(1) R. FIESCHI: *Suppl. Nuovo Cimento*, **1**, 1 (1955).

(2) S. R. DE GROOT: *Thermodynamics of Irreversible Processes* (Amsterdam and New York, 1951).

(3) H. B. CALLEN: *Phys. Rev.*, **73**, 1349 (1948), **85**, 16 (1952); R. FIESCHI, S. R. DE GROOT and P. MAZUR: *Physica*, **20**, 259 (1954).

(4) The symbols we use here have been introduced and defined in I.

and the corresponding change of the conjugated thermodynamical « forces » are written in the more usual form.

Let us write the entropy flow $\boldsymbol{J}'_s$ and the electronic flow $\boldsymbol{J}'_e$ with respect to the lattice:

$$\boldsymbol{J}'_s = \boldsymbol{J}_s + s\varrho(\boldsymbol{v} - \boldsymbol{v}_i)\,, \tag{1}$$

$$\boldsymbol{J}'_e = \boldsymbol{J}_e + \varrho_e(\boldsymbol{v} - \boldsymbol{v}_i) = \frac{\varrho}{\varrho_i}\boldsymbol{J}_e\,, \tag{2}$$

where use has been made of the relations $\boldsymbol{J}_e = -\boldsymbol{J}_i = -\varrho_i(\boldsymbol{v}_i - \boldsymbol{v})$ and $\varrho = \varrho_i + \varrho_e$. $\boldsymbol{v}$ is the barycentric velocity, $\boldsymbol{v}_i$ the velocity of the ionic lattice, ϱ the total density, ϱ_e and ϱ_i the densities of electrons and of ions, and s the entropy per unit mass.

As is well known the « forces » $\boldsymbol{X}'_k$ conjugated with these new fluxes can be found as linear combinations of the former forces $\boldsymbol{X}_k$, by noting that the expression for the entropy production σ is invariant under the transformation from the former to the new set of thermodynamical quantities:

$$-T\sigma = \boldsymbol{J}_e\cdot\boldsymbol{X}_e + \boldsymbol{J}_s\cdot\boldsymbol{X}_s = \boldsymbol{J}'_e\cdot\boldsymbol{X}'_e + \boldsymbol{J}'_s\cdot\boldsymbol{X}'_s\,, \tag{3}$$

hence

$$\boldsymbol{J}_e\cdot\boldsymbol{X}_e + \boldsymbol{J}_s\cdot\boldsymbol{X}_s = \boldsymbol{J}_s\cdot\boldsymbol{X}'_s + \boldsymbol{J}_e\cdot\frac{\varrho}{\varrho_i}(\boldsymbol{X}'_e + s\boldsymbol{X}'_s)\,, \tag{4}$$

where the expressions (1) and (2) for the new fluxes have been used. Relation (4) is satisfied if we put

$$\boldsymbol{X}'_s = \boldsymbol{X}_s \tag{5}$$

and

$$\boldsymbol{X}'_e = \frac{\varrho_i}{\varrho}\boldsymbol{X}_e - s\boldsymbol{X}_s\,. \tag{6}$$

By sybstituting in (6) the explicit expressions for $\boldsymbol{X}_e$ and $\boldsymbol{X}_s$, namely

$$\boldsymbol{X}_s = \operatorname{grad} T\,, \tag{7}$$

$$\boldsymbol{X}_e = \operatorname{grad}(\tilde{\mu}_e - \tilde{\mu}_i)\,, \tag{8}$$

(the time derivative of the vector potential $\boldsymbol{A}$ has been neglected), one has,

$$\begin{aligned}\boldsymbol{X}'_e &= -s\operatorname{grad} T + \frac{\varrho_i}{\varrho}\operatorname{grad}(\tilde{\mu}_e - \tilde{\mu}_i) \\ &= -s\operatorname{grad} T + \frac{\varrho_i}{\varrho}[(e_e - e_i)\operatorname{grad}\varphi + \operatorname{grad}\mu_e - \operatorname{grad}\mu_i]\,,\end{aligned} \tag{9}$$

where the definition of electrochemical potential has also been used: $\tilde{\mu}_k = \mu_k - e_k\varphi$, μ_k being the chemical potential of component k, e_k the charge density per unit mass and φ the electric potential. The Gibbs-Duhem relation, when pressure, as in our case, is uniform, gives

$$-\operatorname{grad}\mu_i = \frac{\varrho_e}{\varrho_i}\operatorname{grad}\mu_e + \frac{s\varrho}{\varrho_i}\operatorname{grad} T. \tag{10}$$

By substituting this expression for $\operatorname{grad}\mu_i$ in (8) we have

$$\boldsymbol{X}'_e = \frac{\varrho_i}{\varrho}(e_e - e_i)\operatorname{grad}\varphi + \operatorname{grad}\mu_e, \tag{11}$$

and finally, as $\varrho_i \simeq \varrho$ and $e_e \gg e_i$ (the charge density per unit mass of electrons is much larger than the charge density per unit mass of ions), and putting $e = e_e$,

$$\boldsymbol{X}'_e = e\operatorname{grad}\varphi + \operatorname{grad}\mu_e. \tag{12}$$

The entropy production can now be written

$$\sigma = -T^{-1}\left(\boldsymbol{J}'_s\cdot\operatorname{grad} T + \frac{\boldsymbol{i}}{e}\cdot\operatorname{grad}\tilde{\mu}_e\right), \tag{13}$$

where the electric current density $\boldsymbol{i}' = e\boldsymbol{J}'_e$ has been introduced. Assuming, according to the scheme of the Onsager theory, linear relations between the « fluxes » and « forces » which appear in (12), we have the phenomenological equations in the desired form:

$$\boldsymbol{J}'_s = -\mathbf{L}_{ss}\cdot\operatorname{grad} T - e^{-1}\mathbf{L}_{se}\cdot\operatorname{grad}\tilde{\mu}_e, \tag{14}$$

$$\boldsymbol{i}' = -\mathbf{L}_{es}\cdot\operatorname{grad} T - e^{-1}\mathbf{L}_{ss}\cdot\operatorname{grad}\tilde{\mu}_e. \tag{15}$$

Among the phenomenological coefficients of equations (14) and (15) the reciprocal relations which have been proved in I still hold, since the validity of these relations is conserved by linear transformations of fluxes and forces which leave the entropy production invariant (2).

* * *

Thanks are due to Dr. P. Mazur for useful discussions.

SUPPLEMENTO
AL VOLUME II, SERIE X, DEL
NUOVO CIMENTO
A CURA DELLA SOCIETÀ ITALIANA DI FISICA

1955 2° Semestre N. 5

Lingua e stile di Galileo (*).

T. Bolelli
Istituto di Glottologia dell'Università - Pisa

(ricevuto il 14 Ottobre 1955)

Uno dei giudizi più comunemente ripetuti su Galileo scrittore fino alla critica crociana è che egli sia stato il creatore della prosa scientifica italiana. È ovvio che Galileo è il creatore soltanto della sua propria prosa (1) e che, se il suo esempio è stato di incalcolabile valore per fondare una tradizione, l'uso del volgare faceva parte di quella rivoluzione scientifica da cui nacque l'indagine moderna e veniva, per così dire, suggerito dalle stesse circostanze dell'insegnamento e della polemica.

La portata immensa delle scoperte di Galileo, il bisogno di comunicare ad una cerchia vastissima di scolari e di studiosi i frutti delle sue osservazioni e delle sue meditazioni sono ragioni che bastano a spiegare l'uso del volgare invece del tradizionale latino. Vi è, però, anche un'altra ragione, che va cercata nell'intimo dello spirito e della formazione stessa di Galileo uomo e scrittore: il bisogno di chiarire a se stesso e agli altri ogni aspetto della ricerca, la necessità di non lasciare nessuna zona dell'indagine nell'ombra e di render possibile il dialogo a tutti, quasi bisogno imperioso di misurare sulle altrui capacità di ragionamento e di comprensione ciò che la mente gli aveva sug-

(*) Conferenza tenuta a Pisa alla Domus Galileana il 10 Maggio 1954.
(1) V. U. Bosco: *Galileo scrittore*, in *La Cultura*, pp. 110-118 (1932).

gerito. Il 16 Agosto 1636 (XVI, p. 475 [2]) scriveva a Fulgenzio Micanzio:

> « Non mancherò di far provisione di tutto il resto delle mie opere per mandargliele [a L. Elzevir] e, se sarà possibile, tutte latine: se ben per ver dire, dove oltre alle serrate dimostrazioni pure matematiche entrano discorsi, nel trasportar l'opere dalla lingua del loro autore in un'altra, si perde assai di grazia, e forse di energia e anco di chiarezza ».

Non sembra che Galileo indugiasse mai troppo a considerare che il latino era richiesto dal pubblico dei dotti stranieri, così come Roberto Galilei gli faceva notare da Lione (XVI, p. 72) a proposito delle *Nuove Scienze;* anzi, apertamente e senza esitazione, dà egli stesso ragione del suo scrivere in volgare (XI, p. 335):

> « Mi è convenuto scriver questo Discorso [è il *Discorso intorno alle cose che stanno in su l'acqua*, del 1612] in lingua italiana, acciò possa esser inteso, almeno in parte, da tutta la città, perchè così ha portato l'occasione di certa disputa ».

Ma vi è di più. Come ha fatto notare il Migliorini [3], pare che Galileo, sia pure per un solo istante, abbia sperato che la lingua italiana potesse diventare la lingua internazionale della scienza. Nella sua terza lettera delle macchie solari dice di aver scritto in italiano anche perché Marco Welser, tedesco di Augusta, già studente a Padova e buon conoscitore della nostra lingua, risponda in italiano (V, p. 189):

> « Dispiacemi ancora della difficoltà che apporta ad Apelle l'aver io scritto nella nostra favella fiorentina: il che ho fatto per diversi rispetti, uno dei quali è il volere in certo modo abusare la ricchezza e perfezion di tal lingua, bastevole a trattare e spiegar e' concetti di tutte le facoltadi; e però dalle nostre Accademie e da tutta la città vien gradito lo scrivere più in questo che in altro idioma. Ma in oltre ci ho auto un altro mio particolar interesse, ed è il non privarmi delle risposte di V. S. in tal lingua, vedute da me e da gli emici miei con molto maggior diletto e meraviglia che se fossero scritte del più purgato stile latino; e parci nel leggere lettere di locuzione tanto propria, che Firenze estenda i suoi confini, anzi il recinto delle sue mura, sino in Augusta ».

In un primo tempo Galileo aveva scritto: « sino in Germania », poi corresse per un innato senso di prudenza e di misura. Tali speranze, sia pure appena accennate di Galileo, avevano un fondamento. Siamo nel 1612 e, se anche

---

[2] Cito dai volumi dell'Edizione nazionale delle Opere di Galileo.

[3] In *Galileo e la lingua italiana*. A questo lavoro, pubblicato nel vol. *Lingua e cultura* (Roma, 1948), pp. 135-158, dobbiamo molti suggerimenti.

qualche tentativo di dotti stranieri era stato fatto per abbandonare il latino (da BACONE nel 1605, quando uscì in inglese *Of Proficience and Advancement of Learning*, subito, però, riscritto in latino col titolo *De dignitate et augmentis scientiarum* o, prima ancora, da CHEPLERO che, nel 1599, pubblicò un almanacco in tedesco per agevolare ai nobili e ai prelati la lettura e, corrispondentemente, a se stesso un congruo aumento di stipendio), la partita era ancora pienamente aperta. CARTESIO solo nel 1637 faceva uscire il *Discours de la méthode* in cui si legge:

> « Si j'écris en français, qui est la langue de mon pays, plutôt qu'en latin, qui est celle de mes précepteurs, c'est à cause que j'espère que ceux qui ne se servent que de leur raison naturelle toute pure jugeront mieux de mes opinions que ceux qui ne croient qu'aux livres anciens. Et pour ceux qui joignent le bon sens avec l'étude, lesquels seuls je souhaite pour mes juges, ils ne seront point, je m'assure, si partiaux, pour le latin, qu'ils refusent d'entendre mes raisons, pour ce que je les explique en langue vulgaire ».

Esigenze simili a quelle di GALILEO sono dunque affacciate almeno venticinque anni più tardi da CARTESIO che, tuttavia, anche in seguito scriverà in latino traducendo poi in francese.

GALILEO rimase fedele, dopo il 1610, e cioè dopo il *Sidereus Nuncius*, al volgare e se anche nelle sue opere non troviamo più nessun accenno ad una qualunque possibilità di vedere l'italiano diventare la lingua universale della scienza, il suo persistere nel volgare è pur sempre significativo indizio di segreta speranza.

Il punto su cui GALILEO insiste è la necessità che ciascuno si renda personalmente conto degli argomenti trattati ed abbandoni il dogmatismo nella scienza. In una gustosissima lettera del 16 Giugno 1612 (XI, p. 327) diretta al GUALDO, egli scrive:

> « Ho ricevuto dal S. Velsero avviso come la mia gl'è pervenuta e che gl'è stata grata; ma che Apelle per hora non potrà vederla, per non intender la lingua. Io l'ho scritta vulgare perché ho bisogno che ogni persona la possi leggere, e per questo medesimo rispetto ho scritto nel medesimo idioma questo ultimo mio trattatello [il *Discorso sulle Galleggianti*]; e la ragione che mi muove, è il vedere, che mandandosi per gli Studii indifferentemente i gioveni per farsi medici, filosofi, etc., sì come molti si applicano a tali professioni essendovi inettissimi, così altri, che sariano atti, restano occupati o nelle cure familiari o in altre occupazioni aliene dalla litteratura, li quali poi, benché, come dice Ruzzante, forniti d'un *bon snaturale*, tutta via, non potendo vedere le cose scritte in *baos*, si vanno persuadendo che in que' *slibrazzon ghe suppie de gran noelle de luorica e de filuorica, e conse purassé che strapasse in elto purassé*; et io voglio ch'e'

vegghino che la natura, sì come gl'ha dato gl'occhi per veder l'opere sue così bene come a i *filuorichi*, gli ha anco dato il cervello da poterle intendere e capire. Contutto ciò vorrei che anco Apelle e gl'altri oltramontani potessero vederla; e qui, per esser io occupatissimo, haverei bisogno del favore di V. S. e del S. Sandeli, il quale mi facesse grazia di trasferirla quanto prima in latino e mandarmela poi subito, perché in Roma è chi si è preso cura di farla stampare insieme con alcune altre mie ».

Che, in verità, prima di Galileo le scienze in Italia non fossero state trattate in volgare, non si può affermare. Luca Pacioli e il Tartaglia basterebbero a mostrare il contrario; non vi è dubbio, tuttavia, che né l'uno né l'altro furono apprezzati per quanto il primo fosse toscano. Erano rimasti, per questo rispetto, degli isolati. Contro di loro vi erano la tradizione scolastica e la scienza ufficiale con tutto il loro peso. Le matematiche, per quanto ritenevano di filosofico e non si limitassero a semplice computo, restavano pur sempre nella sfera della tradizione dotta.

La rivoluzione — perché tale è veramente — dell'uso del volgare avrebbe potuto essere anticipata di almeno un secolo se le opere di Leonardo fossero state sistematiche e fossero diventate note ad un pubblico vasto.

Nella vita di Galileo scritta da Niccolò Gherardini si legge (XIX, p. 642):

« Questo gran personaggio [Gustavo re di Svezia], intrattenendosi per alcuni mesi in Padova, si trovò quasi sempre presente alle lezioni del Sr. Galileo, il quale a richiesta di quei gran Signori fu persuaso a leggere in lingua toscana; conciossiacosaché essendo tutti questi assai ben instrutti nella latina favella et in qualche parte dirozzati nell'italiana, desideravano nell'istesso tempo ch'imparavano le scienze et acquistarne d'essa la perfezzione: e da indi in poi di rado e quasi mai fu udito il S. Galileo con altra lingua fuori che con la natia, etiamdio nella pubblica cattedra, favellare; il che diede materia ad alcuni suoi poco amorevoli di tacciarlo come di poco pratico nell'idioma latino: ma la verità è che ciò facea per compiacere alla voglia degli scolari, la maggior parte oltramontani, e per mettere in reputazione il parlar toscano, con adattare acconciamente i termini d'esso alle conclusioni di filosofia e di mathematica, senza dimandargli in presto o mendicargli da altra lingua che non fosse la propria, contro l'oppenione dei più, quali per addietro ciò stimavano inconvenevole, anzi impossibile ».

Nelle postille del Viviani si legge, però, a correzione del Gherardini:

« Non è vero che leggesse in toscano in pubblico, ma al più qualche lezzione straordinaria »;

e la notizia è ripetuta nel *Racconto istorico della vita di Galileo* dello stesso VIVIANI (XIX, p. 629):

> « E volle [re GUSTAVO] nell'istessa casa di lui (con l'interesse d'esercitarsi insieme nelle vaghezze della lingua toscana) sentire l'esplicazione della sfera, le fortificazioni, la prospettiva e l'uso di alcuni strumenti geometrici e militari, con applicazione et assiduità di vero discepolo, discoprendogli in fine con amplissimi doni quella regia maestà ch'egli s'era proposto di occultare ».

Stando, dunque, al VIVIANI, GALILEO non fece lezione pubblica in volgare: tanto poterono anche su di lui le tradizioni accademiche, alle quali, come il MIGLIORINI ha fatto notare, si era eccezionalmente sottratto, nella prima metà del '500 FRANCESCO VERINO che, secondo le colorite parole del GELLI (4° Ragionamento dei *Capricci del Bottaio*) « leggendo [nello studio] filosofia e veggendo talvolta venire a udirlo il capitano Pepe, il quale non intendeva la lingua latina, subito cominciava a leggere in vulgare perché e' potesse intendere anch'egli. E di poi, poco innanzi che egli si morisse [1543], per dimostrare la inestimabile bontà sua, leggendo pubblicamente ne lo Studio fiorentino il duodecimo libro de la divina Filosofia d'Aristotele, volse esporlo in vulgare acciocché ogni qualità di uomo lo potesse intendere; affermando insieme con Paulo Apostolo di esser così debitore a gli indotti come ai dotti ».

Queste parole sono rivolte dall'anima a Giusto il bottaio, che le obietta:

> « Ma coteste cose di filosofia possonsi elleno dire in vulgare? ».

E l'anima:

> « O perché no? non è la lingua vulgare così ben atta a manifestare i concetti suoi come la latina, e l'altre che son tenute belle e buone? ».

E, ai rinnovati dubbi di Giusto, che si era appellato ai dotti:

> « Giusto, questa è una di quelle cose che la fa dir loro la invidia; ma ei non ci andrà molto tempo, mercé del nostro illustrissimo Duca, che, seguitando di esaltarla, come egli ha incominciato, vi saranno levati questi vetri gialli da gli occhi, che vi fanno vedere ogni cosa giallo ».

La profezia doveva realizzarsi; solo, invece che del Duca, il merito sarebbe stato di GALILEO.

Il *Sidereus Nuncius*, del 1610, segna l'addio di GALILEO al latino, ma precedenti volgari vi erano già stati con *La bilancetta* del 1586 e con le prime lettere del 1588 in cui lo stile appare maturo e sia pure meno sciolto che nelle opere maggiori. Anche molte minute furono stese in volgare per lezioni tenute in latino.

Fra le opere latine di GALILEO occorre, naturalmente, distinguere ciò che è compendio di lezioni altrui, come per esempio i riassunti delle lezioni di un maestro pisano sul *De coelo* di ARISTOTELE (I, p. 24) e le opere originali, i *Theoremata* (1585), il *De motu* (1590), il *Sidereus Nuncius* (1610).

Nei riassunti i valori di stile sono minimi e le qualità usuali ed impersonali: nelle opere originali, il fervore della scoperta, il lievito di pensiero, l'evolversi dell'indagine danno un accento particolare alla lingua che assume un aspetto personale nell'impasto dei volgarismi e delle forme del parlato, pur nella solennità del periodare latino (1). Come è da aspettarsi, negli scritti latini vi sono elementi volgari e nelle opere volgari del primo periodo fino a *La bilancetta* vi sono dei giri di frase latineggianti. Vediamo qualche esempio. Nei *Theoremata*, si legge (I, p. 30):

> « Quaeritur, 4°, an mundus sit perfectus. Videtur non esse: tum quia in illo reperiuntur monstra et defectus, tum quia non semper habet suas species actu existentes, tum quia illi possunt addi multae species perfectiores; quae omnia imperfectionem declarant. Adverte, pro solutione, aliquid posse dici perfectum dupliciter: vel simpliciter, vel secundum quid. Perfectum simpliciter, dicitur illud cui nihil addi potest in perfectione; perfectum secundum quid, illud est quod in suo quidem genere est perfectum, non tamen habet perfectionem omnem possibilem ».

È lo stile solito dei compendi, degli studi scolastici con una, per così dire, nervosa sollecitazione volgare. Il contenuto è ancora tradizionale.

Diversamente nel *De motu*, in cui GALILEO già si allontana dalla dottrina aristotelica ed esprime ormai una materia in cui è impegnato direttamente (I, p. 334):

> « Oh, quam facile ex veris principiis verae extrahuntur demonstrationes! Si verum est, quod diximus, mobilia, dum a statu quietis recedunt, cum tanta contraria vi impressa recedere, quanta est eorum gravitas; ergo quae graviora fuerint, cum maiori virtute contraria coniuncta recedent: quod si gravioribus plus de virtute ad contrarias partes impellente est absumendum quam levioribus, graviora certe ut tardius moveantur necesse erit, cum maiorem patiantur resistentiam. Et si, rursus, haec vera sunt, sequitur ut graviora, postquam tantum de resistentia contraria absumpserint ut non amplius tanto impediantur quanto leviora, ocius descendere debeant: quod certe rursus experientia monstrat ».

---

(1) Si vedano le belle pagine di R. SPONGANO in *La prosa di Galileo e altri scritti*, Messina-Firenze, 1949, p. 93 e seg.

L'impegno, che abbiamo visto espresso con quell'esclamazione all'inizio del periodo or ora citato, si acuisce quando la polemica diventa diretta (I, p. 302):

> « Aristotelem parum in geometria fuisse versatum, multis in locis suae philosophiae apparet; sed in hoc potissimum, ubi asserit, motum circularem motui recto non esse proportionatum, quia, scilicet, recta linea curvae non est proportionata aut comparabilis: quod quidem mendacium (indignum enim est nomine opinionis), nedum intima et magis recondita geometriae inventa, Aristotelem ignorasse, verum et minima etiam principia huius scientiae, demonstrat. Nanque, quomodo dixit rectam et circularem non essere proportionatas, si ad hoc, ut quantitates inter se proportionem habeant, sufficit ut minor possit toties multiplicari ut alteram excedat? An forsan chorda arcus, quae arcu minor est, in se saepius multiplicata, arcus longitudinem non excedet? At si excedet, cur dicitur ab Aristotele, arcum et chordam non esse proportionata? ».

Questo mosso discorso che spesso ricorre ad interrogazioni retoriche e ad improvvisi arresti, questo inseguirsi di idee nuove si accompagna alla libertà del lessico pieno di elementi volgari trasportati in latino: *bilanx*, *ad differentiam*, *peripheria*, *ad minus*, *media* « mezzi », ecc. (5).

Nel *Sidereus Nuncius* il tono di acuta polemica lascia il campo a descrizioni ampie e in qualche punto commosse (III, p. 59):

> « Magnum sane est, supra numerosam inerrantium Stellarum multitudinem, quae naturali facultate in hunc usque diem conspici potuerunt, alias innumeras superaddere oculisque palam exponere, antehac conspectas numquam, et quae veteres ac notas plusquam supra decuplam multiplicitatem superent.
>
> Pulcherrimum atque visu iucundissimum est, lunare corpus, per sex denas fere terrestres semidiametros a nobis remotum, tam ex propinquo intueri, ac si per duas tantum easdem dimensiones distaret ».

In queste parole sembra palpitare la commozione dei nuovi orizzonti aperti alla visione dell'uomo. Tale accento si avverte anche là dove la descrizione si fa più minuta e più rigorosamente scientifica (III, p. 62):

> « De facie autem Lunae, quae ad aspectum nostrum vergit, primo loco dicamus. Quam, facilioris intelligentiae gratia, in duas partes distinguo, alteram nempe clariorem, obscuriorem alteram: clarior videtur totum hemisphaerium ambire atque perfundere, obscurior vero, veluti nubes quaedam, faciem ipsam inficit maculosamque reddit. Istae autem

(5) V. R. Spongano: op. cit., p. 102.

maculae, subobscurae et satis amplae, unicuique sunt obviae, illasque aevum omne conspexit; quapropter magnas, seu antiquas, eas appellabimus, ad differentiam aliarum macularum amplitudine minorum, at frequentia ita consitarum, ut totam Lunarem superficiem, praesertim vero lucidiorem partem, conspergant; hae vero a nemine ante nos observatae fuerunt ».

Già abbiamo accennato alle movenze latine delle prime opere volgari. Ecco come incomincia *La bilancetta* (I, p. 215):

« Sì come è assai noto a chi di leggere gli antichi scrittori *cura si prende, avere Archimede trovato* il furto dell'orefice nella corona d'oro di Ierone, così parmi esser stato sinora ignoto il modo che sì grand'uomo *usar dovesse* in tale ritrovamento; atteso che il credere che procedesse, come da alcuni è scritto, co 'l mettere tal corona dentro a l'aqqua, avendovi prima posto altrettanto di oro purissimo e di argento separati, e che dalle differenze del far più o meno ricrescere o traboccare l'aqqua venisse in cognizione della mistione dell'oro con l'argento, di che tal corona *era composta*, par cosa, per così dirla, molto grossa e lontana dall'esquisitezza ».

Il periodo, con la sua paludata e retorica complessità, arieggia in tutto un periodo latino: ci sono gli infiniti con l'accusativo, i verbi principali sono posti in fondo e questo procedimento d'inversione assurge quasi a sistema. Se Galileo avesse costantemente seguita questa via non sarebbe certo da annoverare fra i maggiori scrittori italiani.

Ad ogni modo, già nella sua esperienza di scrittore latino e nelle sue prime scritture volgari, Galileo si mostra aderentissimo alle cose e, se non proprio immune da retorica, soltanto ai margini di essa. La parola è in dipendenza diretta del pensiero e lo rappresenta con immediatezza. Il pensiero non cede di fronte alla lingua ma la informa arditamente di sé.

Ma è ora di affrontare il quesito fondamentale che si pone parlando di Galileo scrittore. In che cosa consiste l'importanza di Galileo nella nostra storia letteraria? Le opinioni dei critici sono discordi. Abbandonato col Bosco il mito di Galileo creatore della prosa scientifica italiana, vi sono stati finora due modi di intendere Galileo scrittore: o col Sapegno [6], come continuatore della tradizione cinquecentesca, o, con lo Spongano [7], come un'apparizione nuova, originale, senza nulla in comune con gli scrittori del Cinquecento, se si esclude il mezzo di espressione, la lingua toscana. Queste due diverse posizioni, validamente sostenute dai due illustri critici, si fondano su un problema che va chiarito.

---

[6] In *Galileo scrittore* (*Atti e Mem. Accad. dell'Arcadia*, 1948).
[7] Op. cit.

Parlare di Cinquecento in forma astratta, in generale, come se costituisse una categoria assoluta, non pare legittimo. Vi è un Cinquecento del Machiavelli e del Guicciardini e un Cinquecento del Bembo e del Castiglione: vi è un Cinquecento dell'Aretino e del Doni e un Cinquecento di Leonardo e del Cellini. Vi sono nel Cinquecento scrittori sinceri ed immediati, ve ne sono altri immersi nella retorica, e non soltanto in quella di stile elevato, ma anche in quella (meno osservata ma non meno reale) popolaresca. Il Cinquecento ha una corrente latineggiante e boccaccevole che si sviluppa fuori di Toscana ed una corrente più fresca, propria di autori toscani, tanto che un giovane studioso ha potuto recentemente parlare [8] di un edonismo linguistico del Cinquecento.

Considerando partitamente gli scrittori del Cinquecento sarebbe molto difficile stabilire una filiazione di Galileo; se, d'altra parte, ci riportiamo ad ideali di misura e di chiarezza, se facciamo questione di gusto e di decoro e confrontiamo il suo stile con quello di Bruno e di Campanella, allora i legami col Cinquecento appaiono innegabili. In questo caso si comprende come la pagina notissima del Viviani, citata da tutti i critici, abbia sapore e valore anche per il critico attuale per scoprire la poetica di Galileo (XIX, p. 627):

> « Fu dotato dalla natura d'esquisita memoria; e gustando in estremo la poesia, aveva a mente, tra gl'autori latini, gran parte di Vergilio, d'Ovidio, Orazio e di Seneca, e tra i toscani quasi tutto'l Petrarca, tutte le rime del Berni e poco meno che tutto il poema di Lodovico Ariosto, che fu sempre il suo autor favorito e celebrato sopra gl'altri poeti, avendogli intorno fatte particolari osservazioni e paralleli col Tasso sopra moltissimi luoghi.... Parlava dell'Ariosto con varie sentenze di stima e e d'ammirazione; et essendo ricercato del suo parere sopra i due poemi dell'Ariosto e del Tasso, sfuggiva prima le comparazioni, come odiose, ma poi, necessitato a rispondere, diceva che gli pareva più bello il Tasso, ma che gli piaceva più l'Ariosto, soggiugnendo che quel diceva parole, e questi cose. E quand'altri gli celebrava la chiarezza et evidenza nell'opere sue, rispondeva con modestia, che se tal parte in quelle si ritrovava, la riconosceva totalmente dalle replicate letture di quel poema, scorgendo in esso una prerogativa solo propria del buono, cioè che quante volte lo rileggeva, sempre maggiori vi scopriva le meraviglie e le perfezioni ».

Lo scrittore prediletto di Galileo è, dunque, un poeta, l'Ariosto, e questo suo orientamento è rivelatore. La congenialità del grande scienziato col poeta del Rinascimento, è fondata su elementi di gusto, di quel gusto che poneva,

---

[8] C. Segre in *Giorn. st. lett. ital.*, **70**, 1953, p. 145, sgg.

fra Galileo e il Tasso, una barriera di incomprensione, che faceva tacere nel giudizio critico di Galileo i lati positivi del poeta della *Gerusalemme* coi suoi inquieti chiaroscuri, col corrispondersi continuo delle passioni intime col mondo reale, col contrappunto di nuove melodie e risonanze ignote al classicismo del primo Cinquecento.

Di chiarezza e di evidenza si parla nel Viviani: è lo spirito fiorentino contro una forma di decadentismo che tende verso morbidezze a Galileo ignote o sgradite. Nella sua avversione per gli « arzigogoli », gli « scambietti metafisicali », le « capriole intrecciate » del Tasso, in questa sua pedanteria nella valutazione di un pur grande poeta, troviamo l'orientamento sicuro per penetrare nello studio dello stile galileiano.

Non sarà neppur male notare che, in qualche caso, egli dà sulla voce persino al prediletto Ariosto, là dove in minime giunture, in rari accostamenti di parole gli pare poco spontaneo, poco chiaro e alquanto duro. Ad esempio « per fuggire la durezza del secondo verso e la continuazione di otto i » propone (IX, p. 185) di mutare il verso ariostesco (*Orlando Furioso*, XXXVII, 82):

« Da le mogli così furo i mariti
da le madri così i figli divisi »

in questo modo:

« E così dalle mogli i lor mariti
e i figli dalle madri fur divisi ».

Qui pare di sentire Galileo musico insorgere un po' pedantescamente contro una cacofonia dell'amato ed ammirato poeta. E non vi è dubbio che, fra i caratteri estrinseci della prosa galileiana, vi è una accurata sapienza stilistica, una chiarezza, una sostenuta armonia, un tono sempre identicamente dignitoso pur nella varietà degli atteggiamenti.

A comprendere la sua prosa, servono anche i suoi errori estetici, come segno di gusto e come presupposto del suo concreto modo di scrivere.

Anche il fatto che Galileo, a detta del Viviani (XIX, p. 625), « era... provvisto di pochissimi libri », aiuta a comprendere la singolarità della sua composizione. Perché, se è vero, come dice il Momigliano che « la storia generale della nostra prosa è anzitutto storia di classicismo e di retorica » [9], è anche vero che alcuni fra i nostri autori più grandi furono immuni dalla retorica e retorico non fu Galilei [10]. A tenerlo lontano dalla retorica fu la sua mente scientifica, il suo restare a contatto continuo con la realtà, il suo ingenito senso della misura delle cose.

---

[9] *Elzeviri*, pp. 50-56.

[10] Una ricerca degli elementi barocchi nella prosa di Galilei, ad opera di Giulio Marzot in *Convivium* (1954, pp. 678-689 e 1955, p. 43), così si conclude: « Noi riteniamo che il Galilei sia da dirsi scrittore di evidenza barocca: ma barocco solo per quel tanto che la classicità, vale a dire l'armonia dei suoi spiriti e modi, lo consentiva ».

Lo spirito di osservazione si tramuta in lui, come dice il LEOPARDI, in « scolpitezza evidente » ([11]), giudizio, sia pure accomunante GALILEO col REDI, che riscatta quell'ambigua osservazione (« Il nostro Galileo lo chiami elegante chi non conosce la nostra lingua »), che si rifà ad una infelice frase del GIORDANI che era venuto fuori, parlando dell'« acutissimo ed elegantissimo ingegno » del Cardinale PALLAVICINO, con queste parole: « Al sommo Galileo sovrabbondò la mente, ma parve quasi mancare lo studio nell'opera di scrivere; in quella copia diffuso e soverchio, talora languido, talora confuso » ([12]). Ma nello stesso LEOPARDI vi è un altro giudizio, contrastante singolarmente con questo: « Di questa associazione della precisione coll'eleganza, è splendido esempio lo stile di Celso e fra' nostri di Galileo » ([13]).

Da tale contraddizione è possibile arguire l'incertezza del giudizio critico per la singolarità stessa della prosa galileiana. La quale sorge, non già su una tradizione formata (almeno per quanto ogni prosa si rifà a una scuola), ma su una nuova visione del mondo e accompagna fedelmente le speculazioni scientifiche e filosofiche con larghezza e chiarezza di impostazione e con una singolare capacità di mantenersi sempre ad un livello elevatissimo. Egli diceva che non si possono unire « i fioretti retorici con la saldezza delle dimostrazioni » ed aggiungeva (VI, p. 330):

> « Io non so vedere perché, potendo uno dir bene assolutamente con una semplicissima e propriissima parola, ei debba servirsi d'una impropria e bisognosa di limitazioni ed insomma d'esser finalmente trasportata in un'altra molto diversa ».

Si è detto che il ragionamento del MACHIAVELLI procede secondo una forma di catena; altrettanto avviene del ragionamento di GALILEO in cui si notano, però, anelli solidamente ribattuti. I suoi periodi sono lunghissimi e la più genuina ed immediata impressione che si ha leggendoli è come di meraviglia perché quella complessità di ragionamento non stanca mai: tutto si regge su un'armonia e su un equilibrio che ha quasi il valore di una forza della natura. Le esemplificazioni si possono trovare in ogni pagina del *Saggiatore*, dei *Massimi sistemi*, delle *Nuove scienze*. Leggiamo questo passo del *Saggiatore* (VI, p. 251):

> « ... Qualunque volta sia presente l'effetto, necessariamente vi è anco quella causa. E così, per darne un esempio, chi dicesse, " Il tale ha acceso il fuoco, adunque si è servito dello specchio ustorio ", errerebbe, potendo derivar l'accendimento dal batter un ferro, dall'esca e fucile, dalla confricazione di due legni, e da altre cause; ma chi dicesse " Io ho sentito batter il fuoco al vicino ", e soggiungesse " Adunque egli ha della pietra

([11]) *Zibaldone*, ediz. curata da F. FLORA, I, p. 45.
([12]) Op. cit., II, p. 65 e la nota a p. 1342.
([13]) Op. cit., I, p. 886.

focaia ", senza ragione sarebbe ripreso da ch i gli opponesse che, concorrendo a tale operazione, oltre alla pietra, il fucile, l'esca e 'l solfanello ancora, non si poteva con buona logica inferir la pietra risolutamente. E così, se l'ingrandimento dell'angolo e l'union de' raggi concorrono sempre nell'operazioni del telescopio, delle quali una è il far veder l'invisibile, perchè da questo effetto non si può inferire quale delle due cause più ne piace? Io credo di penetrare in parte la mente del Sarsi, il quale, s'io non m'inganno, vorrebbe che il lettore credesse quello ch'egli stesso assolutamente non crede, cioè ch'il veder le stelle, che prima erano invisibili, derivasse non dall'ingrandimento dell'angolo, ma dall'unione de' raggi; sì che, non perchè la specie di quelle divenisse maggiore, ma perchè i raggi fussero fortificati, si facesser visibili: ma non si è voluto apertamente scoprire, perchè troppo gli sono addosso l'altre ragioni del Sig. Mario taciute da esso, ed in particolare quella del vedersi gl'intervalli tra stella e stella ampliati colla medesima proporzione che gli oggetti quaggiù bassi; i quali intervalli non dovrian ricrescer punto, se niente ricrescessono le stelle, essendo loro così distanti da noi come quelle ».

È difficile, credo, trovare uniti nello stesso periodo, come avviene in GALILEO ad ogni pagina, un'onda di proposizioni così ampia e nello stesso tempo così chiara. Il discorso è pieno, non lascia sottintendere nulla, ogni cosa è espressa con pari minuziosa cura, tutto è incatenato sapientemente e la lettura procede con quel senso di compiutezza e quasi di dovizia che non cade mai nell'eccessivo. Il freno è dato da quel sentimento del limite che deriva dal gusto dell'autore e dall'insegnamento delle cose dalle quali germoglia una lezione di misura e di proporzione.

Nella prosa di GALILEO, anche le sentenze sono di per sè mirabili. Splendida a me sembra la seguente (VI, p. 340):

« Se il discorrere circa un problema difficile fusse come il portar pesi, dove molti cavalli porteranno più sacca di grano che un caval solo, io acconsentirei che i molti discorsi facessero più che un solo; ma il discorrere è come il correre,e non come il portare, ed un caval berbero solo correrà più di cento frisoni ».

Tuttavia quasi sempre esse perdono di vigore e di sapore se sono staccate dal contesto perchè, in generale, si incorporano indissolubilmente al tessuto (VI, p. 237):

« Ma pur fussero i veri filosofi come l'aquile, e non più tosto come la fenice! Signor Sarsi, infinita è la turba degli sciocchi, cioè di quelli che non sanno nulla; assai son quelli che sanno pochissimo di filosofia; pochi son quelli che ne sanno qualche piccola cosetta; pochissimi quelli che ne sanno qualche particella; un solo Dio è quello che la sa tutta. Sì che,

per dir quel ch'io voglio inferire, trattando della scienza che per via di dimostrazione e di discorso umano si può dagli uomini conseguire, io tengo per fermo che quanto più essa participerà di perfezzione, tanto minor numero di conclusioni prometterà d'insegnare, tanto minor numero ne dimostrerà, ed in conseguenza tanto meno alletterà, e tanto minore sarà il numero de' suoi seguaci ».

Il rivolgersi direttamente all'avversario che vuol confutare è uno dei procedimenti più frequenti nello stile di GALILEO e si trova non solo nelle opere maggiori come il *Saggiatore* ma anche in quelle annotazioni marginali e in quelle postille che aprono la via a comprendere i ragionamenti di maggior impegno.

La polemica e l'ironia si accompagnano spesso a queste movenze stilistiche e talvolta lo sdegno prorompe e il sarcasmo giunge ben a fondo nelle carni dell'oppositore. Celebre è quel passo contro ANTONIO ROCCO, nelle note marginali alle *Esercitazioni filosofiche* (VII, p. 270):

« Deh non avessi io mai scoperte queste novità in cielo, di tante innumerevoli non prima vedute stelle fisse, di quel che siano le nebulose, la Via Lattea, de' collaterali di Saturno, della corte di Giove, la immensa mutazione di grandezza in Marte, le importune macchie nel Sole, le grandi mutazioni di figura e grandezza in Venere, le scabrosità grandissime nella Luna; deh mai io non l'avessi palesate al mondo poi che dovevano concitarmi l'odio del signor Antonio Rocco e di tanti altri signori filosofi! Consolatevi, Signore; ché il tempo, scopritore della verità, in breve è per estirpare queste fallacie, e più le vane conseguenze che io stoltamente ne deducevo, e i vostri scritti, pieni di dottrina ferma e soda, viveranno immortali, ad onta delle mie esorbitantissime chimere ».

Lo stile di GALILEO, sempre tenuto su un piano altissimo, si accende, nell'urto polemico immediato, fino al sarcasmo e alla passionalità che arriva all'ingiuria. A proposito di un'osservazione di LUDOVICO DELLE COLOMBE (autore di un discorso contro il moto della terra), GALILEO annota (III, p. 258):

« Sarebbero matti [i copernicani] a rispondere a queste pippionate ».

L'ironia e la malizia sorgono sempre quando il pacato ragionare non ha fatto presa su menti ostinate a non voler capire e su occhi che si rifiutano di vedere, come nel caso di quel peripatetico che, assistendo ad una operazione anatomica e constatando che i nervi si originano nel cervello, dice (VII, p. 134):

« Voi mi avete fatto veder questa cosa talmente aperta e sensata, che quando il testo di Aristotile non fusse in contrario, che apertamente dice, i nervi nascer dal cuore, bisognerebbe per forza confessarla per vera ».

O altrove (III, p. 398):

> « Tu di' che ne i sensibili comuni non si deve credere a un senso senza l'attestazione dell'altro; adunque tu vuoi creder più a due testimoni falsi che a un solo »;

in cui, quel rivolgersi direttamente all'interlocutore fa venire in mente le improvvise apostrofi di Leonardo quando interpella un presunto lettore:

> « Nissuna umana investigazione si pò dimandare vera scienzia, s'essa non passa per le matematiche dimostrazioni e se *tu dirai* che le scienzie, che principiano e finiscono nella mente, abbiano verità, questo non si concede... ».

La prosa di Galileo è così il frutto di una mente di singolari capacità creative in cui la concatenazione logica non porta a monotonia o a scorie di struttura, in cui il tono commosso di fronte alle nuove visioni della natura si unisce alla ferma obiettività del ragionamento e la integra.

Il culmine letterario della prosa galileiana è il *Saggiatore*, dove il calore della polemica e l'evidenza delle immagini, il procedimento logico e l'evasione poetica, il passare dalla sentenza al motto di spirito, l'incalzare l'avversario e l'abbandonarlo per inseguire una verità improvvisamente balenata danno una tale varietà alla pagina da porre la grandezza di Galileo scrittore accanto a quella di Galileo scienziato. Fin dalle prime pagine tali qualità sono evidenti. In quella sentenza (VI, p 214):

> « il contraddire alla geometria è un negare scopertamente la verità ».

par di sentire ancora una volta Leonardo con le sue lodi alle « matematiche dimostrazioni » e il suo anelito verso la verità.

Lo stesso piglio di alcune locuzioni: « lasciando il suo scolare, come si dice, nelle peste », oppure: « spero d'aver a fare uscir voglia ad alcuno di molestare (come si dice) il mastino che dorme, e voler briga con chi si tace » con quelle attenuazioni in proposizioni incidentali, dà una nota di colore all'inizio del trattato e sembra il sorriso di chi sa di avere già vinto. Così nel motteggiare sul titolo dell'opera del Sarsi (*Libra astronomica ac philosophica*), dice (VI, p. 220):

> « Ma perchè m'è paruto che, nel ponderare egli le proposizioni del Sig. Guiducci, si sia servito d'una stadera un poco troppo grossa, io ho voluto servirmi d'una bilancia da saggiatori, che sono così esatte che tirano a meno d'un sessantesimo di grano ».

Ma nella fine del capitolo il periodo diventa più severamente rilevato per una osservazione del Sarsi a proposito dello Scorpione (VI, p. 221):

> « Adunque molto più proporzionatamente, ed anco più veridicamente, se riguarderemo la sua scrittura stessa, l'avrebbe egli potuta intitolare:

L'Astronomico e filosofico Scorpione, costellazione dal nostro sovran poeta Dante chiamata figura del freddo animale " che colla coda percuote la gente "; e veramente non vi mancano punture contro di me, e tanto più gravi di quelle degli scorpioni, quanto questi, come amici dell'uomo, non feriscono se prima non vengono offesi e provocati, e quello morde me che mai nè pur col pensiero non lo molestai. Ma mia ventura, che so l'antidoto e rimedio presentaneo a cotali punture! Infragnerò dunque e stropiccerò l'istesso scorpione sopra le ferite, onde il veleno risorbito dal proprio cadavero lasci me libero e sano ».

In questo periodo l'evidente realismo dell'immagine ha una potenza dantesca. Un'apertura solenne assume il discorso quando si parla della natura (VI, p. 232):

« Parmi, oltre a ciò, di scorgere nel Sarsi ferma credenza, che nel filosofare sia necessario appoggiarsi all'opinioni di qualche celebre autore, sì che la mente nostra, quando non si maritasse col discorso d'un altro, ne dovesse in tutto rimanere sterile ed infeconda; e forse stima che la filosofia sia un libro e una fantasia d'un uomo, come l'*Iliade* e l'*Orlando Furioso*, libri ne' quali la meno importante cosa è che quello che vi è scritto sia vero. Sig. Sarsi, la cosa non istà così. La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscere i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto ».

Nei *Massimi sistemi* il procedimento dialettico assume forma dialogica, il che permette una maggior varietà stilistica e la costruzione di caratteri, di personaggi che s'imprimono nella memoria del lettore con una loro netta individualità.

L'osservazione dei fenomeni naturali si amplia in un orizzonte più vasto, in temi più grandi che portano a scrutare l'universo nella sua miracolosa immensità.

In qualche caso l'aggettivazione si fa ricca, complessa, vuole accompagnare la dimostrazione in ogni particolare anche minimo perché nulla sfugga all'occhio esperto dell'osservatore (VII, p. 132):

« Per quello dunque che mi è restato in mente, fu la somma dei discorsi di ieri l'andar esaminando da i fondamenti loro, qual delle due opinioni sia più probabile e ragionevole: quella che tiene, la sustanza de i corpi celesti esser ingenerabile, incorruttibile, inalterabile, impassibile, ed in somma esente da ogni mutazione, fuor che dalla locale, e però essere

una quinta essenza diversissima da questa de i nostri corpi elementari, generabili, corruttibili, alterabili, etc. ».

Lo stupore per le opere della natura si esprime talvolta con una serie di interrogazioni che lasciano come un'eco profonda, uno strascico di risonanze e di armonie (VII, p. 128):

> « Non direm noi che 'l sapere scoprire in un marmo una bellissima statua ha sublimato l'ingegno del Buonarruoti assai sopra gli ingegni comuni degli altri uomini? E questa opera non è altro che imitare una sola attitudine e disposizione di membra esteriore e superficiale d'un uomo immobile; e però che cosa è in comparazione d'un uomo fatto dalla natura, composto di tante membra esterne ed interne, de i tanti muscoli, tendini, nervi, ossa, che servono a i tanti e sì diversi movimenti? Ma che diremo de i sensi, delle potenze dell'anima, e finalmente dell'intendere? non possiamo noi dire, e con ragione, la fabbrica d'una statua cedere d'infinito intervallo alla formazion d'un uomo vivo, anzi anco alla formazion d'un vilissimo verme? ».

Anche qui credo che non si vada errati se si richiama l'ammirato stupore di fronte alla natura di Leonardo da Vinci.

Fra le figure più umanamente vere degli scritti di Galileo spicca Simplicio, attaccato ad una tradizione che non si rassegna a morire, ostinato come don Ferrante e sempre pronto a trovar mille cavilli, ad opporre ragioni, anche se poco valide, alla ricerca di una verità che ormai travolge e trionfa. Il suo procedere sillogistico, che non arriva ad arginare l'incalzante ragionamento dei due interlocutori, talvolta si esprime con alternative che vorrebbero esser conclusive ma non sono; né è raro il caso che nell'affollato suo stile scaturisca qualche forma popolaresca (VII, p. 128):

> « O io non sono un di quegli uomini che intendano, o 'n questo vostro discorso è una manifesta contradizione. Voi tra i maggiori encomii, anzi pur per il massimo di tutti, attribuite all'uomo, fatto dalla natura, questo dell'intendere; e poco fa *dicevi* con Socrate che 'l suo intendere non era nulla; adunque bisognerà dire, che né anco la natura abbia inteso il modo di fare un intelletto che intenda »,

in cui quel *dicevi* attribuito alla seconda persona plurale è una forma eminentemente popolaresca.

Nonostante il loro atteggiamento di battaglia, resta, tuttavia, indubbio che i *Massimi sistemi* ha uno stile più disteso del *Saggiatore*, più ampiamente discorsivo, più calmo e meno nervoso.

In qualche punto viene in mente Platone, come nel passo seguente in

cui un accenno alla natura o all'ora del giorno fa ritornare sulla terra la mente estatica e perduta nella contemplazione dei prodigi dell'ingegno umano (VII, p. 130):

> «Io son molte volte andato meco medesimo considerando, in proposito di questo che di presente dite, quanto grande sia l'acutezza dell'ingegno umano; e mentre io discorro per tante e tanto meravigliose invenzioni trovate da gli uomini, sì nelle arti come nelle lettere, e poi fo reflessione sopra il saper mio, tanto lontano dal potersi promettere non solo di ritrovarne alcuna di nuovo, ma ancora di apprendere delle già ritrovate, confuso dallo stupore ed afflitto dalla disperazione, mi reputo poco meno che infelice. S'io guardo alcuna statua delle eccellenti, dico a me medesimo: E quando sapresti levare il soverchio da un pezzo di marmo, e scoprire sì bella figura che vi era nascosa? Quando mescolare, e distendere sopra una tela o parete colori diversi, e con essi rappresentare tutti gli oggetti visibili, come un Michelangelo, un Raffaello, un Tiziano? S'io guardo quel che hanno ritrovato gli uomini nel compartir gl'intervalli musici, nello stabilir precetti e regole per potergli maneggiar con diletto mirabile dell'udito, quando potrò io finir di stupire? Che dirò de i tanti e sì diversi strumenti? La lettura de i Poeti eccellenti di qual meraviglia riempie chi attentamente considera l'invenzion de' concetti e la spiegatura loro? Che diremo dell'architettura? che dell'arte navigatoria? Ma sopra tutte le invenzioni stupende, qual eminenza di mente fu quella di colui che s'immaginò di trovar modo di comunicare i suoi più reconditi pensieri a qualsivoglia altra persona, benché distante per lunghissimo intervallo di luogo e di tempo? parlare con quelli che sono nell'Indie, parlare a quelli che non sono ancor nati né saranno se non di qua a mille e dieci mila anni? e con qual facilità? con i vari accozzamenti di venti caratteruzzi sopra una carta. Sia questo il sigillo di tutte le ammirande invenzioni umane, e la chiusura de' nostri ragionamenti di questo giorno: ed essendo passate le ore più calde, il signor Salviati penso io che *avrà gusto di andare a godere de i nostri freschi in barca*; e domani vi starò attendendo amendue per continuare i discorsi cominciati».

Nei dialoghi delle *Scienze Nuove*, la tendenza ad un discorso più specificatamente scientifico ha il sopravvento. Anche i personaggi hanno ormai una funzione quasi di pura struttura. Ci avviciniamo al tipo di prosa scientifica moderna che mira a giungere direttamente alle dimostrazioni.

Nelle *Scienze Nuove* abbiamo ancora un periodare armonioso ma manca quella commossa partecipazione umana che faceva così suggestive le pagine del *Saggiatore*.

L'impressione di misura, di proporzione e di armonia che si ha leggendo

le opere di GALILEO è confermata se si considera da vicino il suo atteggiamento nei riguardi della terminologia studiata acutamente dal MIGLIORINI [14]. Uno scienziato fra i più grandi e fra i più fecondi di nuove scoperte, fra i più ricchi di intuizioni come GALILEO non ha introdotto termini nuovi nel vocabolario italiano. Egli accoglie senza ribellarsi le voci greche e latine già entrate nel vocabolario della scienza come *sesquialtero*, *prostaferesi*, *parallasse*, *ipogeo*, *transonoro*, ma ha un profondo ritegno, che par quasi inibizione, nel coniare termini nuovi. Per supplire alle necessità terminologiche delle sue scoperte, GALILEO usa con significato tecnico vocaboli già in uso. È il caso di « pendolo » o di « ancora » (per armare la calamita), di tanto preesistenti a GALILEO, che vengono, per così dire, tecnificati ed adoprati con un nuovo uso scientifico. Spesso si può seguire nelle pagine di GALILEO il passaggio di un termine dal suo valore generale a quello tecnico [15].

Questo atteggiamento di GALILEO porta alla necessità di definire e fissare i termini nel nuovo significato e dà un più umano ed aperto respiro alla sua pagina. Egli non ha mai voluto sacrificare ad una etichetta il largo e sicuro fraseggiare, l'ariosa ampiezza della sua pagina. Avviene, così, che anche le più caratteristiche invenzioni galileiane sono state tenute a battesimo da altri.

Caratteristica è, a questo proposito, la storia della voce *telescopio*. Subito dopo la scoperta, GALILEO chiamò la sua invenzione *cannone* o *occhiale*, poi *cannocchiale*, formazione, in verità, non felice. Nel *Sidereus Nuncius* GALILEO usò *perspicillum*, CHEPLERO sostituì a questo *conspicillum*, il padre SCHEINER (Apelle) *helioscopium*. Fu FEDERICO CESI che, nel 1611 (almeno), introdusse *telescopio*. Da allora anche GALILEO adottò questo termine che divenne poi generale. Il *microscopio*, altra gloria di GALILEO, fu così chiamato da GIOVANNI FABER, accademico linceo di Bamberga, il *termoscopio* deve il suo nome al BIANCANI (1620) e poi divenne *termometro* ad opera del padre LEURECHON; la *cicloide* ebbe il suo nome dal padre MERSENNE.

GALILEO rimase fermo nel suo atteggiamento per tutta la sua attività di scrittore e di scienziato. Quella che egli chiamò *bilancetta* fu chiamata *idrostammo* dagli Accademici del Cimento ed è stato dimostrato [16] che molte voci, che compaiono nella Crusca e nel TOMMASEO come attestate per la prima volta in GALILEO sono dei suoi predecessori: *ascensionale* è già in COSIMO BARTOLI, *gravità* nel TARTAGLIA, *nebulosa* in EGNAZIO DANTI, così come *selinografia* e *selinografo* risalgono a BACONE [17].

Come GALILEO non ha creato termini nuovi, così ha contribuito a seppellirne parecchi: è tutta una massa di errori che viene battuta inesorabilmente.

---

[14] Op. cit., p. 145.
[15] Per « pendolo », vedi *Massimi sistemi* in VII, p. 177 e 256.
[16] Dal MIGLIORINI, op. cit., p. 150.
[17] *Novum Organum*, (1620); II, XXXIX: « *selenographia* ».

Così anche lo studio della terminologia di GALILEO si conclude con una lezione di misura. È, questo, uno dei caratteri che distinguono appunto la lingua di GALILEO da quella dei suoi scolari e seguaci. La grandezza, l'afflato poetico sempre rigorosamente misurato e controllato, diventa, nella prosa scientifica del '600, in VIVIANI, nel CASTELLI, nel MAGALOTTI, nel REDI, nel BELLINI, decoro, e perfino piacevolezza e colore o trapunto d'immagini non disgiunte da accademismo. Resta l'efficacia, la dignitosa chiarezza: non vi è più la magnanima commozione, l'ammirato stupore di chi ha rivelato per primo all'umanità mondi nuovi, lontanissimi orizzonti, di chi ha vissuto tutta la vita nella scoperta e nella contemplazione di nuove verità.

La prosa di GALILEO, sorta in una iniziale simbiosi di latino e volgare, raggiunse, nelle scoperte che egli quotidianamente faceva, una commossa perfezione, una vivacità e una robustezza quali ben di rado si riscontrano nei nostri scrittori. La duttilità del suo mezzo espressivo ci conduce quasi a identificare la sua realtà scientifica con un favoloso mondo in cui nebulose e stelle, leggi fisiche e osservazioni matematiche si vestono di un alone di leggenda. Nelle soste più felici della sua osservazione dell'immensità dei cieli, dello spazio infinito e nella contemplazione degli esseri più piccoli, GALILEO è sempre presente con la sua acutissima mente (XIII, p. 208):

> « Io ho contemplati [col microscopio] moltissimi animalucci con infinita ammirazione: tra i quali la pulce è orribilissima, la zanzara e la tignuola son bellissimi; e con gran contento ho veduto come faccino le mosche et altri animalucci a camminare attaccati a' specchi, et anco di sotto in su.... In somma ci è da contemplare infinitamente la grandezza della natura, e quanto sottilmente ella lavora e con quanta indicibil diligenza ».

Oppure (X, p. 280):

> « Io mi trovo al presente in Venezia per fare stampare alcune osservazioni, le quali col mezo di un mio occhiale ho fatte ne i corpi celesti; et sì come sono di infinito stupore, così infinitamente rendo grazie a Dio, che si sia compiaciuto di far me solo primo osservatore di cosa ammiranda et tenuta a tutti i secoli occulta ».

Poi, per quegli occhi che avevano spaziato nell'infinito, venne la più atroce condanna, la cecità. Nel gennaio 1638 così GALILEO annunziava al DIODATI la sua sventura (XVII, p. 247):

> « Signor mio, il Galileo, vostro caro amico e servitore, è fatto irreparabilmente da un mese in qua del tutto cieco. Or pensi V. S. in quale afflizione io mi ritrovo, mentre che vo considerando che quel cielo, quel mondo e quello universo che io con mie maravigliose osservazioni e chiare

dimostrazioni avevo ampliato per cento e mille volte più del comunemente veduto da' sapienti di tutti i secoli passati, ora per me s'è sì diminuito e ristretto, ch'e' non è maggiore di quel che occupa la persona mia ».

È stato detto che nessuno scienziato dei tempi moderni è al tempo stesso un grande scrittore come GALILEO. Queste parole, così piene di pietà, così contenute nel loro dolore ci fanno ancor più persuasi della grandezza del suo animo e ci avvicinano alla sua tormentata eppur serena umanità senza la quale non si potrebbe spiegare la singolare grandezza del suo stile, l'indimenticabile ed insostituibile valore della sua esperienza di scienziato e di scrittore.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 5, 1955
2° Semestre

Albert Einstein and the Theory of Relativity.

C. LANCZOS

Dublin Institute for Advanced Studies - Dublin

(ricevuto il 17 Dicembre 1955)

> «O frati,» dissi, «che per cento milia
> perigli siete giunti all'occidente,
> a questa tanto picciola vigilia
>
> de' vostri sensi, ch'è del rimanente,
> non vogliate negar l'esperienza,
> *diretro al sol, del mondo sanza gente.*
>
> Considerate la vostra semenza:
> fatti non foste a viver come bruti,
> ma per seguir virtute e canoscenza».
>
> DANTE, *Divina Commedia*,
> Inferno, XXVI, 112-120.

With the death of ALBERT EINSTEIN on April 18, 1955, a world-figure leaves the scene who will be remembered by future historians as the most profound genius of our era and one of the greatest thinkers of all ages. He opened to the human mind new perspectives of unprecedented grandeur and demonstrated the comprehensibility of the physical universe to a degree which was not surmised before.

The present article is not devoted to the man EINSTEIN, to the human personality who was of extraordinary stature in itself. The many popular articles he wrote on urgent ethical, political and philosophical issues of our days [42, 43] show the same earnestness and penetrating power of the mind that characterize his scientific writings. They are, however, outside the scope of our present discussions which are solely devoted to EINSTEIN the physicist. But even as a physicist he was the originator of so many profound discoveries that their due evaluation would require a volume. If we concentrate here on a brief sketch of his scientific achievements, we do that in the hope that the very brevity of the enumeration will give evidence of the astonishing scope of his scientific results.

If one studies the scientific output of ALBERT EINSTEIN, one is immediately struck by the dual strain of his meditations: EINSTEIN, the ingenious statistician, and EINSTEIN, the ingenious geometer. Here are two modes of thinking which are very seldom united in one person's mind to such an outstanding degree. But in the magnificent early investigations of EINSTEIN both strains run parallel and his contributions in both directions were equally profound. While it was the dazzling discovery of general relativity which singled him out to that pedestal that he will occupy forever in the history of scientific and philosophical thought, yet the development of contemporary physics in its quantum aspects was even stronger influenced by his penetrating analyses of the radiation-matter problem. If we deal here more in detail with his relativistic speculations, we do that partly because he himself turned in the second half of his life more and more exclusively to this field of thought, and partly because the final impact of his genius on man's intellectual evolution will in all probability reside in that realm. It would be injust, however, to omit altogether that side of EINSTEIN's work which would have made him one of the leading physicists of his era, even without any of his relativistic discoveries.

EINSTEIN belonged to that group of great scientists who like to submerge in a problem unprejudiced by the opinions of other workers in the field. The mere reading of literature and the mere absorption of scientific results in ready-made form give them little satisfaction; (a habit frequently shared by lesser minds). It is not so surprising that the university student EINSTEIN did not excel to any undue degree. He was not interested in the essentially formalistic mathematical courses to which he was exposed. He was not attracted to the intricacies of mathematical formalism and mathematics for its own sake did not excite his enthusiasm — in marked contrast to his last years, when he was inclined to trust purely formal relations perhaps too strongly. His mind was glued on the physical structure of the universe and he developed an almost uncanny intuition which allowed him to roam freely in the land of physical concepts, deducing new relations by a most sparing use of mathematical tools. EINSTEIN never cared for techniques and never cared for details which had no relation to something fundamental. This put his papers in a special category and was responsible for the peculiar under-estimation which many scientists extended to them. In many scientific treatises of today, and in some historical expositions of contemporary physics, EINSTEIN's role is astonishingly under-rated. In our over-technicized era most scientists become proficient in certain techniques and value technical achievements more than anything else. Achievements of a sweeping nature, obtained by ingenious intuition rather than by systematic drudgery, are suspected as vague and philosophical rather than strictly scientific. But fortunately for EINSTEIN's personal fate, the greatest minds of European physics, particularly PLANCK, VON LAUE and SOMMERFELD in Germany, later RUTHERFORD and EDDINGTON in England,

LANGEVIN and DE BROGLIE in France, LORENTZ in Holland, recognized the extraordinary stature of EINSTEIN's genius and gave him credit where credit was due. The Commemorating Volume published at the occasion of his seventieth birthday in 1949 [44] (1), contained articles from an international galaxy of physicists and the number of dissenters was negligible.

1. – Thermodynamics was a subject which fascinated EINSTEIN particularly, because of its universal validity. Thermodynamics is a science which permits us to establish relations between the various physical and chemical properties of matter on the basis of a few very general principles, without going into the specific mechanisms which are the real originators of these relations. This is made possible by the very large number of essentially identical material particles which cooperate in a statistical way to bring about a certain macroscopically measurable effect. The pyramid is there, while the millions of slaves whose painstaking efforts collaborated in erecting the pyramid, remain invisible. Around the beginning of this century, in the student days of the young EINSTEIN, many physicists doubted the existence of atoms which they considered a figment of scientific imagination, in spite of the great results of the kinetic theory of gases, developed by MAXWELL and later by BOLTZMANN and GIBBS. The methods of statistical mechanics, which dealt with the statistical interaction of a huge number of submicroscopic particles, established a perfect counterpart to classical thermodynamics by showing the atomistic-kinetic background of the thermodynamical principles. But a direct physical proof of the existence of atoms was not available yet. In his first thermodynamical investigations EINSTEIN rediscovered certain fundamental relations which were given earlier by BOLTZMANN and GIBBS [1, 2, 3]. He was not satisfied, however, with the microscopic justification of the macroscopic picture but tried to come to results which would bring the kinetic action of the molecules in direct evidence. In the memorable year 1905, which saw the birth of three of EINSTEIN's great discoveries,—the Brownian motion, the light quanta hypothesis and special relativity—he succeeded with his plans [5]. He applied the statistical theory to a particle suspended in a fluid and found that the irregular heat-action of colliding molecules is the source of a diffusion equation with the diffusion constant

$$D = \frac{RT}{N}\,\frac{1}{6\pi\eta a} \qquad \begin{array}{l}(\eta = \text{viscosity coefficient}\\ a = \text{radius of particle})\,.\end{array}$$

The result of this diffusion equation is that the average displacement of the suspended particle will not be zero but a quantity which increases with the

(1) Cf. also: *In Commemoration of the Seventieth Birthday of Albert Einstein, March 14, 1949*, in *Rev. Mod. Phys.*, **21**, 343-540 (1949).

square root of the time t; (for a time t which is not too small):

$$\lambda_x = \sqrt{\overline{x^2}} = \sqrt{2Dt}\,.$$

Actual measurements of the peculiar trembling motion of suspended particles—called the Brownian motion—fully corroborated Einstein's equation and palpably demonstrated the existence of atoms and molecules, converting to the atomic hypothesis all the physicists who previously took a sceptical stand. But EINSTEIN realized that more is at stake. It was the investigation of Planck's radiation law which stimulated him to the highest flights of his statistical discoveries. He conceived the idea that instead of using the microscopic picture as an explanation of the workings of thermodynamics, one might revert the procedure and draw from the macroscopic thermodynamical relations conclusions concerning the structure of the microscopic mechanism. PLANCK deduced his radiation law from the assumption that energy is not emitted in continuous quantities but in packages which are multiples of $h\nu$. It seemed unnecessary to make a more detailed picture of the radiation mechanism; in particular, one could assume that the Maxwellian equations hold in vacuum and the quantum laws come into operation only in the interaction between matter and radiation. But EINSTEIN inquired into the forces which act on a material particle suspended in a field of radiation and found that a thermodynamical equilibrium between matter and radiation could never come about if the Maxwellian equations would hold exactly. He evaluated the energy fluctuations which must exist in a small volume of the radiation field. These energy fluctuations are of fundamental importance since it is their effect on the material particle which makes the equilibrium between radiation and matter possible. The average value of the energy fluctuations in the frequency range $(\nu, \nu+\mathrm{d}\nu)$ could be calculated on the basis of the following thermodynamical relation [10]:

$$\overline{\varepsilon^2} = kT^2 \left(\frac{\partial E}{\partial T}\right)_V .$$

Planck's energy law gave the result

$$\overline{\varepsilon^2} = h\nu E + \frac{c^3}{8\pi\nu^2 \mathrm{d}\nu}\,\frac{E^2}{V} .$$

The second term is the only one which could be explained on the basis of Maxwell's equations. But its contribution in the realm of high frequencies is by far too small to convey the proper statistical energy to the suspended material particle.

The first term had exactly the same form as the energy fluctuation of a gas with particles of the energy $h\nu$. This resulted in the amazing picture of

a « light quantum » (now called « photon »), as a small corpuscle endowed with the energy $h\nu$ and moving with light velocity c. How this corpuscular picture could be harmonized with the wave picture of radiation, was (and essentially remained) an unsolved riddle but EINSTEIN had the boldness of mind to recognize that the hypothesis has to be made, in spite of its strangeness [4]. He immediately applied it to a theoretical explanation of the photo-electric effect, and later to the high-frequency limit of the continuous spectrum emitted by a fast electron when it is captured.

EINSTEIN's next problem was whether the absorption and re-emission of light energy by an oscillator could be conceived as a purely scalar effect, comparable to the emission of a spherical wave, or whether it had directional qualities. He could show convincingly that both the emission and absorption of light must be conceived as a directional effect by forming the picture of a light quantum as a corpuscle which carries the momentum $h\nu/c$ in the direction of its propagation.

In the meantime he made the first fundamental advance in the field of specific heat by applying Planck's energy law to the elastic oscillations which are present in a solid body. EINSTEIN simplified the picture by assuming one single mode of vibration only, yet he obtained a dependency of the specific heat on the temperature which was correct in its main character and came close to the physical data in the realm of very low temperatures [8] (2).

A very beautiful contribution to Planck's radiation law was obtained by EINSTEIN when under the impact of Bohr's theory he pondered on an interaction mechanism between matter and radiation which is in harmony with the general features of Bohr's atom model [20]. He utilized only the most general properties of Bohr's theory, in particular the assumption that a particle can possess only certain discrete energy levels ε_m. Between two energy levels a certain transition probability A_m^n is assumed by EINSTEIN, i.e. a certain percentage of molecules (per unit time) can spontaneously change over from ε_m to ε_n if $\varepsilon_m > \varepsilon_n$, emitting the energy $\varepsilon_m - \varepsilon_n$. On the other hand a certain transition probability B_n^m exists in the opposite direction, absorbing the corresponding $\varepsilon_m - \varepsilon_n$. This transition requires, of course, the presence of a certain radiation density $\varrho(\nu, T)$; the percentage of particles which participates in the transition will be proportional to $B_n^m \varrho$. But the opposite process, « induced emission », (in addition to spontaneous emission) can take place too and will be proportional to $B_m^n \varrho$. Finally we need the statistical weights p_n and p_m of the two energy states ε_n and ε_m.

(2) The summation procedure over all the elastic vibrations was later carried out by P. DEBYE (*Ann. der Phys.*, **39**, 789 (1912)) and independently by M. BORN and TH. VON KARMAN (*Phys. Zeits.*, **13**, 297 (1912)).

Now the condition of radiation equilibirum gives

$$p_n \exp\left[-\frac{\varepsilon_n}{kT}\right] B_n^m \varrho = p_m \varepsilon \exp\left[-\frac{\varepsilon_m}{kT}\right](B_m^n \varrho + A_m^n),$$

to which we can add the relation

$$p_n B_n^m = p_m B_m^n,$$

which follows from the natural assumption that the radiation density $\varrho(\nu, T)$ should go to infinity as T goes to infinity. From these two fundamental relations we obtain

$$\varrho(\nu, T) = \frac{A_m^n / B_m^n}{\exp[(\varepsilon_m - \varepsilon_n)/kT] - 1}.$$

On the other hand, we know from general thermodynamics (independently of any specific exchange mechanism between radiation and matter) that Wien's displacement law must hold as to the general structure of $\varrho(\nu, T)$:

$$\varrho(\nu, T) = \nu^3 f\left(\frac{\nu}{T}\right).$$

This immediately gives as the only possibilities:

$$A_m^n = \alpha \nu^3 B_m^n$$

and

$$\varepsilon_m - \varepsilon_n = h\nu$$

with the two constants α and h. We have thus obtained from very general considerations Planck's radiation law, plus the fundamental Bohr condition which determines the emitted (or absorbed) frequency between two energy states.

The last ingenious contribution of EINSTEIN to statistical thermodynamics was instigated by an idea of the Indian physicist S. N. BOSE who demonstrated ([3]) how Planck's radiation law could be obtained as a consequence of a corpuscular statistics in which the light quantum particles are considered as indistinguishable from each other. EINSTEIN recognized the analogy between an ideal monoatomic gas and a radiation field [23, 24]. He showed how the classical thermodynamics of an ideal gas corresponded to a radiation law

([3]) *Zeits. f. Phys.*, **26**, 178 (1924); **27**, 384 (1924).

of the Wien type which violated the Nernst theorem. The new statistics corrected this behavior to a thermodynamical law which corresponded to Planck's law and which removed the contradiction to the Nernst theorem [24]. The « gas degeneracy » thus discovered by EINSTEIN became fundamental for the understanding of the statistical properties of free electrons inside of a conducting solid body.

Somewhat aside from these statistical considerations and yet of great historical significance was EINSTEIN's formulation of the Sommerfeld-Wilson quantum conditions in terms of Hamilton's S-function [19]. The accidental feature of the quantum conditions for separable systems was thus replaced by an invariant formulation which could be applied to non-separable systems too. Einstein's general formulation of the quantum conditions in the form of one unified and invariant principle was interpreted by DE BROGLIE as a *resonance condition* for waves. Hence the DE BROGLIE matter waves take their origin from Einstein's invariant quantum condition. De Broglie's matter waves, on the other hand, culminated in Schrödinger's wave-mechanics. Hence we see the direct historical sequence: EINSTEIN-DE BROGLIE-SCHRÖDINGER.

2. – The discovery of analytical geometry by DESCARTES was a landmark in the history of mathematical thought. It made it possible to reduce geometry to algebra and solve all problems of geometry by purely symbolic tools. The « coordinates » are a pair of numbers x, y (in space a triplet of numbers x, y, z) which are coordinated to a point P and which can replace the point P for all operational purposes.

A student of analytical geometry quickly masters the notion of Cartesian coordinates and finds nothing mysterious in them. EINSTEIN wondered about many things that other people took for granted and this was his strength. He was asked once how it happened that he and not somebody else discovered relativity, His answer was that other people wondered about the properties of space in their early youth but in their mature years saw no problems anymore. He, however, came in touch with the problems of space at a much later age and therefore could ask questions that other people overlooked [47, p. 84].

The question of coordinates disturbed EINSTEIN since his early youth. Coordinates are a wonderful tool in the description of physical events. But why is it that the role of these coordinates is so very different in pure geometry and in physics? In order to define our coordinates we first have to establish a definite set of *axes* to which these coordinates are related. These axes do not belong to the given geometrical problem but can be chosen in any orientation we like. Any time we want we can abandon one set of coordinates in favor of another set. Each set of coordinates establishes a definite « reference system » but all these reference systems have equal validity since space knows no particular center and no preferential directions. No reference

system can be designated as an « absolute system »; on the contrary, all frames of reference are equivalent relative to each other and of equal a priori validity for the description of geometrical relations; (notwithstanding the fact that in a *special* problem we may choose our coordinate axes in a way which is particularly suited to the symmetry properties of the given problem).

In physics the situation is very different. We may observe the very same event from many different reference systems but the description will by no means be equivalent. For example we observe a stone falling to the earth and we describe its motion as one of uniform acceleration. But let us assume that another observer is in a frame of reference which falls with the stone. He will say that the stone is constantly at rest. Its velocity does not increase uniformly with the time t but is permanently zero.

Newton was aware of the difficulty but his opinion was that we have no right to quarrel with creation on philosophical grounds. He assumed as a matter of fact the existence of a certain « absolute frame of reference » in which we have to describe the laws of nature. If we depart from this frame, the fundamental laws of nature will not come out properly.

Actually the absoluteness of the frame of reference was not so extreme in Newton's physics. If we are in a moving train and that train moves on absolutely smooth tracks and with absolutely constant velocity, there is no instrument in the world by which the velocity of the train as such could be measured. If the curtains are drawn and we make no comparisons with the outer world, everything will happen in the moving train exactly as if it were in rest. Hence all reference systems which move with constant velocity relative to each other are equally justified for the description of the laws of nature. The equivalence of reference systems is violated only if *accelerated* systems are taken into account.

In Newton's time the optical and electromagnetic phenomena were not yet properly investigated. When during the nineteenth century Faraday and Maxwell established the field theory of electromagnetism and the identity of optical and electromagnetic phenomena was demonstrated, the problem of the absolute reference system reappeared to an even more menacing degree. In the realm of optical and electromagnetic phenomena an absolute reference system had to be assumed; the previous leeway of reference systems moving relative to each other with constant velocity had to be dropped. Indeed, light is a wave motion and these waves propagate in every direction with the same constant velocity c. But if the observer moves relative to the light waves, he disturbs the homogeneity of light propagation in various directions and this must lead to observable effects by which his motion relative to the absolute reference system (also called « aether ») may be detected. Yet all the manifold efforts to demonstrate these effects failed. The famous Michelson-Morley experiment of 1887 was the first of a long line of distinguished experiments

which used indicators of excessive accuracy for the demonstration of the absolute motion of the earth through the aether but which failed to detect any trace of such motion. Everything occurred exactly in the way as if the earth were in rest—in accordance with that relativity principle which in the realm of purely mechanical phenomena was well established since NEWTON's time.

The tremendous puzzle posed by these negative results led to several attempts of a theoretical explanation, before EINSTEIN entered the scene. G. F. FITZGERALD and H. A. LORENTZ conceived independently the idea that the length of a rod is influenced by its absolute motion in such a way that the ellipsoidal wave measured out with the help of the moving rod appears spherical. The weakness of all these explanations was that they could not make it plausible why nature is constructed in such a way that it prevents us by all means from discovering the absolute motion through the aether.

EINSTEIN approached the problem from a fundamental angle [6]. He conceived the result of the Michelson-Morley experiment and all related experiments as a clear demonstration of the fact that all frames of reference which move relative to each other with uniform velocity, must be equivalent for the description of natural events. Hence the principle of relativity, previously established for the realm of mechanical phenomena only, must be extended to *all* phenomena of nature. But then the question has to be investigated, what relations must exist between various equivalent reference systems. Let x, y, z, t be the coordinates used by one observer and x', y', z', t' the coordinates used by another observer who is in relative motion to the previous observer. It was always our custom to put intuitively

$$t = t' ,$$

i.e. to assume that time is an absolute, universal variable which has validity for the entire universe. It turns out, however, that this cannot be so; we must insist that the homogeneous propagation of light in every direction is equally valid for all observers who are in equivalent reference systems. This has the consequence that the equation of uniform light propagation

$$x^2 + y^2 + z^2 - c^2t^2 = 0$$

must remain valid in the new reference system, if it was valid in the original system. This leads to the relation

$$x^2 + y^2 + z^2 - c^2t^2 = k(x'^2 + y'^2 + z'^2 - c^2t'^2) .$$

Although k is not necessarily 1, yet the natural demand that the transformation formulae shall be linear, together with the condition that the trans-

formation shall be reversible without any preference for the one or the other system, permits us to normalize k to 1. Thus we obtain the principle that the relations between equivalent reference systems are those linear relations between the coordintaes which leave the quadratic form

$$x^2 + y^2 + z^2 - c^2t^2$$

invariant. Such transformations occurred earlier in consequence of certain mathematical properties of the Maxwellian equations and were investigated by LORENTZ and by POINCARÉ. But it was EINSTEIN who discovered the proper interpretation of these transformations as relations between coordinates which belong to equivalent frames of reference. Thus it was shown that time has no absolute significance but is merely a fourth coordinate, to be added to the three space coordinates.

EINSTEIN drew all the fundamental physical consequences from his discovery. The Newtonian equations of motion, and in fact all equations of physics had to be revised, in order to bring them in harmony with the relativity principle. If a certain equation was given in a certain frame of reference—e.g. the frame in which the particle was in momentary rest—one could immediately transform it in a unique way to any other equivalent frame. The new equations of motion showed a definite dependence of the mass m on the velocity, according to the law

$$m = \frac{m_0}{\sqrt{1 - (v^2/c^2)}}.$$

The mass became infinite for light velocity which was the highest attainable velocity under any circumstances since even an arbitrarily large force acting for an arbitrarily long time could never accelerate a particle beyond light velocity. Measurements on fast particles corroborated the consequences of relativistic mechanics beyond any shade of doubt.

If an observer B had the velocity u relative to an observer A and an event occurred relative to B with the velocity v, the velocity of this event relative to A was not $u+v$ but

$$w = \frac{u + v}{1 + (uv/c^2)}.$$

This is EINSTEIN's famous addition theorem of velocities. We can see from it that if v is light velocity, w remains light velocity, no matter how large u may be. In relativity theory light velocity plays the same role as an infinite velocity plays in ordinary mechanics.

In the year 1907 [9] EINSTEIN drew the most important consequence of his theory, viz. the equivalence of mass and energy:

$$E_0 = m_0 c^2 .$$

It expresses the principle that every mass by its very existence is the seat of a tremendous amount of energy which under ordinary circumstances remains dormant, due to the high stability of m_0. In certain nuclear processes, however, involving fission or fusion, a transformation of mass into radiation takes place which means the release of an enormous amount of energy. The terrible destructive power of the Atombomb—which converted only a few ounces of matter into heat, creating an artificial sun of several billion centigrades—verified Einstein's equation in a tragic way.

The converse effect, viz. that every energy has an equivalent mass value, was deduced by EINSTEIN still in 1905 [7]. He showed that a body which receives the energy E in any form whatsoever, must increase its mass according to the equation

$$\Delta m = \frac{E}{c^2} .$$

In 1908 HERMANN MINKOWSKI made a very fundamental mathematical discovery by giving a metrical interpretation of Einstein's theory (4). He showed that the reason for the constancy of light velocity is not a physical but a geometrical fact. It is a direct consequence of the circumstance that the world of physics has a metrical structure which includes space and time into one unified manifold. The square of the distance element of this manifold is not the ordinary Euclidean expression $dx^2 + dy^2 + dz^2$ but the extended expression

$$-\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - c^2\,\mathrm{d}t^2$$

The equations of physics are such that they comply with the homogeneity of a four-dimensional instead of a three-dimensional manifold. Space and time merge into one united manifold, the role of time being an additional fourth dimension. The equivalent reference systems of EINSTEIN are in geometrical interpretation various orientations of the fundamental set of four axes, x, y, z, ict, which can be freely rotated in the four-dimensional world. The inertial systems of Einstein appear in geometrical formulation as the group of four-dimensional rotations which leave the fundamental line-element invariant.

This was a constructive new principle of fundamental importance which

(4) *Math. Annalen*, **68**, 472 (1910).

served EINSTEIN as a departing point for his later investigations. He freely admitted his indebtedness to MINKOWSKI's geometrical discovery and thus the claim that he did not give sufficient credit to MINKOWSKI's work, is hardly justified. EINSTEIN was not a mathematician, of course, and it took him some time before he recognized the fundamental advance achieved by the new geometrical interpretation. But after that he plunged himself whole-heartedly into a further exploration of the new geometry.

One of the important moves of MINKOWSKI which played an important part in the later development of the theory, was the construction of the so-called « matter tensor » T_{ik}. The theory of relativity united the conservation laws of momentum and energy into the four components of one single vector equation. The question arose, however, how these conservation laws should be formulated if energy (or mass) was not lumped in singly particles but distributed continuously over the field. The Mawxellian electromagnetic field gave a model for such a distribution and also the answer to the problem. The Maxwellian stress-components $t_{ik} = t_{ki}$ $(i, k = 1, 2, 3)$ formed a symmetric tensor of second order. They had to be extended to a symmetric tensor of four dimensions. Hence a fourth row t_{4i} and a fourth column t_{i4} had to be added. In physical interpretation t_{44} had the significance of energy density, while the three components t_{i4} $(i = 1, 2, 3)$ represented the momentum density. The three components t_{4i} $(i = 1, 2, 3)$, on the other hand, represented the energy flux. Already POINCARÉ postulated in 1900 that every energy flux must be associated with a corresponding momentum density. Now the equality of energy flux and momentum density became the consequence of the symmetry of the matter tensor.

Quite generally, momentum and energy distributed over a field had to be conceived as the components of a symmetric tensor of second order $T_{ik} = T_{ki}$. The conservation of momentum and energy was expressed by the mathematical equation that the divergences of this tensor vanished:

$$\frac{\partial T_{i\alpha}}{\partial x_\alpha} = 0 .$$

3. Relentlessly the demon drove EINSTEIN to newer and newer efforts. The principle of relativity established the equivalence of reference systems, but so far on a very restricted scale only. Why should *uniform* motions be preferred? Why not *all* motions? It should be possible to formulate the laws of nature in such a way that no reference system would be a priori distinguished. The fundamental laws of nature should be equally describable in *all* reference systems. This was the principle of general relativity which offered a tremendous challenge to EINSTEIN but he knew that he will find no peace before he arrives at the solution of this problem. And here started that dogged up-hill fight

SOCIETÀ ITALIANA DI FISICA

SCUOLA INTERNAZIONALE DI FISICA

Sotto gli auspici del Ministero della Pubblica Istruzione
e del Consiglio Nazionale delle Ricerche

4° Corso estivo a Varenna, sul Lago di Como

15 Luglio - 4 Agosto 1956

1. Quest'anno 1956, col contributo del Ministero della Pubblica Istruzione, del Consiglio Nazionale delle Ricerche, del Comitato Nazionale per le Ricerche Nucleari, dell'Università di Milano, delle Autorità, di Enti, Società e privati della Provincia di Como, la Scuola terrà a Varenna, nella Villa Monastero, messa gentilmente a disposizione dall'Ente Villa Monastero, il *4° Corso estivo*.
   Il Corso sarà dedicato alle *proprietà magnetiche della materia*. In particolare verranno organicamente trattati i seguenti argomenti:
   — *Paramagnetismo nei cristalli*;
   — *Rilassamento paramagnetico*;
   — *Ferromagnetismo, ferrimagnetismo, antiferromagnetismo*;
   — *Risonanza paramagnetica, ferromagnetica, antiferromagnetica e diamagnetica*;
   — *Risonanza magnetica nucleare nei liquidi e nei cristalli*;
   — *Larghezza della riga nella risonanza magnetica*;
   — *Il concetto di temperatura nei fenomeni magnetici*;
   — *Magnetismo a basse temperature, polarizzazione nucleare*;
   — *Proprietà magnetiche dei superconduttori.*
   Le lezioni verranno affidate ai seguenti professori:
   C. J. Gorter, Kamerlingh Onnes Laboratorium der Rijksuniversiteit, Leiden.
   C. Kittel, University of California, Berkeley, California (U.S.A.).
   N. Kurti, Clarendon Laboratory, Oxford.
   L. Néel, Laboratoire d'Électrostatique et de Physique du Métal, Grenoble.
   M. H. L. Pryce, H. H. Wills Physical Laboratory, Bristol.
   E. M. Purcell, Harvard University, Cambridge, Massachusetts (U.S.A.).
   J. H. Van Vleck, Harvard University, Cambridge, Massachusetts (U.S.A.).
   Altri specialisti terranno seminari e conferenze su argomenti connessi con quelli che sono l'oggetto del Corso.
   Le lezioni, i seminari, le conferenze e le discussioni saranno generalmente tenute in inglese o in francese.
2. L'organizzazione e la Direzione del Corso sono affidate al prof. Luigi Giulotto, professore di Fisica Superiore all'Università di Pavia.
3. La durata del Corso sarà di 21 giorni. L'inaugurazione avrà luogo Domenica 15 Luglio alle ore 18 e la chiusura nel pomeriggio di Sabato 4 Agosto.
4. Il numero totale degli allievi sarà di 35.
   Chi desidera partecipare al Corso dovrà inviare entro il 31 Maggio 1956 la domanda al

Prof. LUIGI GIULOTTO
Istituto di Fisica dell'Università
PAVIA (Italia)

fornendo, ben chiaramente, le seguenti indicazioni:

a) Nome e Cognome.
b) Data e luogo di nascita.
c) Indirizzo attuale.
d) Titoli di studio universitari ottenuti e presso quale Università.
e) Professione attuale.
f) Elenco delle pubblicazioni di Fisica.
g) Quale conoscenza orale e scritta ha dell'inglese e del francese.
h) Se verrà a Varenna da solo o accompagnato da familiari e, nel caso che venga da solo, se è disposto a prendere alloggio in camera in comune con altro allievo.

Inoltre, il richiedente deve allegare alla domanda una lettera di presentazione, rilasciata da un professore universitario di Fisica, il quale assicuri che il richiedente ha interesse per i suoi studi a frequentare il Corso e che possiede già adeguata preparazione.

L'accettazione delle domande è decisa dalla Presidenza della Società unitamente alla Direzione del Corso, tenendo conto dei documenti presentati dai singoli richiedenti e di una equa ripartizione dei posti tra le varie nazioni cui questi appartengono.

Le comunicazioni delle deliberazioni saranno inviate agli interessati entro il 10 Giugno 1956.

5. Gli allievi saranno alloggiati o nella foresteria della Villa Monastero o in alberghi di Varenna, vicino alla Villa, in camere da 1 o 2 letti. Nella stessa Villa o in un albergo di Varenna sarà organizzata la mensa della Scuola.

6. Grazie alle sovvenzioni di cui la Scuola dispone, la spesa totale che ogni allievo deve sostenere per la frequenza al Corso, alloggio e vitto alla mensa della Scuola, è limitata a L. 30000, se egli prende alloggio in camera da solo, e a L. 25000, se prende alloggio in camera con altro allievo o con familiari. Dette somme sono da versarsi, non più tardi del 23 Luglio, all'Amministrazione della Scuola, a Varenna, in valuta italiana.

   Borse di studio, in numero molto limitato, potranno essere accordate a quegli allievi che ne giustifichino il bisogno.

7. La Scuola cercherà, nei limiti delle possibilità locali (che sono però assai limitate causa la stagione balneare), di trovare una sistemazione adeguata in alberghi di Varenna per i familiari che intendessero accompagnare gli allievi.

   I familiari possono essere accolti, insieme con l'allievo, alla mensa della Scuola.

   Tutte le spese per alloggio, vitto, ecc., relative ai familiari sono da computarsi a parte e totalmente a carico dell'allievo: esse, per trattamento uguale a quello fatto all'allievo, si aggirano intorno a L. 3000 per giorno e per persona adulta, e dovranno essere regolate direttamente dall'allievo con l'albergatore.

8. Per ogni informazione relativa al Corso rivolgersi direttamente al Direttore di questo.

Milano, 25 Febbraio 1956

Il Segretario della S.I.F.
G. C. Dalla Noce

Il Direttore del Corso
L. Giulotto

Il Presidente della S.I.F.
G. Polvani

ITALIAN PHYSICAL SOCIETY

INTERNATIONAL SCHOOL OF PHYSICS

Under the auspices of the Ministero della Pubblica Istruzione
and the Consiglio Nazionale delle Ricerche

4th Summer Course to be held in Varenna, Lake Como

15th July - 4th August 1956

1. This year, 1956, thanks to contributions from the Ministero della Pubblica Istruzione, the Consiglio Nazionale delle Ricerche, the Comitato Nazionale per le Ricerche Nucleari, the University of Milan, and Authorities, Companies and private individuals of the Province of Como, the *4-th Summer Course* will be held at Varenna in the Villa Monastero, the use of which has been kindly granted by the Ente Villa Monastero.
   The Course will be devoted to the *magnetic properties of matter*. In particular the following subjects will be organically treated:
   — *Paramagnetism in crystals*;
   — *Paramagnetic relaxation*;
   — *Ferromagnetism, ferrimagnetism and antiferromagnetism*;
   — *Paramagnetic, ferromagnetic, antiferromagnetic and cyclotron resonance*;
   — *Nuclear magnetic resonance in liquids and in crystals*;
   — *Line-widths in magnetic resonance*;
   — *The concept of temperature in magnetic phenomena*;
   — *Magnetism at low temperatures, nuclear polarisation*;
   — *Magnetic properties of superconductors.*
   The lectures will be given by following Professors:
   C. J. Gorter, Kamerlingh Onnes Laboratorium der Rijksuniversiteit, Leiden.
   C. Kittel, University of California, Berkeley, California (U.S.A.).
   N. Kurti, Clarendon Laboratory, Oxford.
   L. Néel, Laboratoire d'Électrostatique et de Physique du Métal, Grenoble.
   M. H. Pryce, H. H. Wills Physical Laboratory, Bristol.
   E. M. Purcell, Harvard University, Cambridge, Massachusetts (U.S.A.).
   J. H. Van Vleck, Harvard University, Cambridge, Massachusetts (U.S.A.).
   Other specialists will hold seminars and lectures on questions related to the basic subjects of the course.
   The lectures, seminars, conferences and discussions will be held generally in English or French.
2. The person intrusted with the organisation and Direction of the Course is Mr. Luigi Giulotto, Professor at the University of Pavia.
3. The Course will last 21 days. It will open at 6.00 p.m. on Sunday the 15-th July and will close on the afternoon of Saturday the 4-th August.
4. The total number of students will be 35.
   Those wishing to attend the course shoulds end not later than 31-st of May 1956 an application to:

Prof. LUIGI GIULOTTO
Istituto di Fisica dell'Università
PAVIA (Italy)

and should furnish the following information, which should be set out clearly and legibly:

a) Christian name and surname.
b) Date and place of birth.
c) Present address.
d) Degrees and other qualifications obtained, with name of university in each case.
e) Present professional activity.
f) List of publication in physics.
g) Standard of knowledge of English and French – written and spoken.
h) Whether intending to stay at Varenna unaccompanied or with members of family and in the former case, whether willing to share a room with other students.

The applicant should send, together with his application, a note of reference from an university professor of Physics, to testify for the applicant's interest and preparation in the activities of the School.

Each application will be considered by the President of the Italian Physical Society and by the Direction of the School on the basis of the information submitted, with regard also to a fair distribution of the places available, among students of various Nations. The decisions on the admittance to the School will be made known to the applicants by 10th June 1956.

5. Accomodation for the students will be provided in rooms with 1 or 2 beds either in the guest quarters of the Villa Monastero or in hotels at Varenna, near the Villa. Meals will be taken at the villa itself or at a hotel in Varenna.

6. Thanks to the grants made available to the School the total fee (for attendance, hotel accomodation and full board) will be 30000 Lire for each student having single hotel accomodation, and 25000 Lire for each student sharing hotel accodation with other students or with members of his family. Fees should be paid to the management of the School, at Varenna, in Italian currency, not later than the 23rd of July.

A very limited number of scholarships may be granted to those students whose economic conditions might otherwise prevent them from attending the School.

7. The School will do everything possible to find suitable accomodation in local hotels for members of families accompanying the students. It should be noted, however, that, in view of the holiday season, local possibilities are limited.

Members of student's families may avail themselves of the catering arrangements being organised for the School.

All expenditure involved in the hotel accommodation, board etc., for student's relatives will be payable separately. These expenses, covering full service similar to that enjoyed by the student, will be of about 3000 Lire per adult person per day and are to be settled directly with the hotel management.

8. For all further information regarding the School, application should be made directly to Professor Luigi Giulotto.

Milan, the 25-th February 1956

The Secretary of the Society
G. C. Dalla Noce

The Director of the Summer Course
L. Giulotto

The President of the Society
G. Polvani

ONORANZE AD AMEDEO AVOGADRO

CONGRESSO INTERNAZIONALE SULLE COSTANTI FONDAMENTALI

E

XLII CONGRESSO DELLA SOCIETÀ ITALIANA DI FISICA

TORINO, 6-11 SETTEMBRE 1956

In occasione del centenario della morte di AMEDEO AVOGADRO, Torino, sua città natale, organizza quest'anno un ciclo di manifestazioni per onorare degnamente il grande scienziato.

Nel quadro di queste manifestazioni, la Società Italiana di Fisica organizzerà, sotto gli auspici dell'Unione Internazionale di Fisica Pura ed Applicata e del Consiglio Nazionale delle Ricerche, un Congresso Internazionale di Fisica sulle « costanti fondamentali », che avrà luogo a Torino dal 6 all'11 Settembre 1956. Intendiamo comprendere, sotto questa denominazione generale, non soltanto quelle grandezze fisiche che sono comunemente chiamate « costanti universali », ma anche le grandezze della fisica atomica e nucleare, quali i pesi atomici dei nuclidi, le costanti delle forze nucleari, ecc.

Pertanto il Congresso non tratterà soltanto il problema particolare dei valori più probabili da attribuire alle costanti fondamentali, ma — più generalmente — le questioni connesse con quei rami della Fisica sperimentale e teorica ai quali si può in qualche modo assegnare, come caratteristica, questa o quella costante. E precisamente:

— per la costante di Avogrado: **questioni relative alle molecole e ai cristalli**;

— per la velocità della luce: **questioni relative a particelle molto veloci**;

— per la carica elettrica elementare, la costante di Planck e le costanti delle forze nucleari: **questioni relative ai fenomeni elettrodinamici, alle interazioni mesone-nucleone e alla Fisica dei neutrini**;

— e finalmente, per le costanti caratteristiche dei nuclidi: **questioni relativi ai pesi atomici, ai processi nucleari, ecc.**

Il Congresso sarà evidentemente diviso in diverse Sezioni, che verranno precisate più tardi.

Inoltre, contemporaneamente al Congresso Internazionale, avrà luogo a Torino nella sua forma consueta il XLII Congresso della Società Italiana di Fisica.

Segretario Generale di tutta l'organizzazione è il prof. R. DEAGLIO, Istituto di Fisica dell'Università di Torino (via P. Giuria 1, Torino), al quale coloro che desiderano partecipare al Congresso possono rivolgersi per informazioni, non più tardi del 30 Aprile 1956.

CELEBRATION OF AMEDEO AVOGADRO

INTERNATIONAL CONGRESS ON FUNDAMENTAL CONSTANTS

AND

XLII CONGRESS OF THE ITALIAN PHYSICAL SOCIETY

TURIN, 6th-11th SEPTEMBER, 1956

On occasion of the centenary, 1956, of the death of AMEDEO AVOGADRO, the city of Turin, his birth-place, is going to organize celebrations in honour of the great scientist.

The Italian Physical Society takes part in these celebrations organizing, under the auspices of the International Union of Pure and Applied Physics and of the National Research Council, an International Congress on Fundamental Constants of Physics, to be held in Turin from September 6th to 11th 1955. We wish to include, under this general denomination, not only the physical quantities which are generally called « universal constants », but also the quantities of atomic and nuclear physics such as the atomic weights of nuclides, the constants of nuclear forces, etc.

The Congress will therefore treat not only the particular problem of the most probable values which are to be attributed to the fundamental constants, but more generally, the questions concerning those branches of experimental and theoretical physics to which one can attach, as a characteristic, this or that constant. And exactly:

— for the constant of Avogadro: **questions concerning molecules and crystals**;

— for the velocity of light: **questions concerning very fast particles**;

— for the elementary electric charge, the constant of Planck and the constants of nuclear forces: **questions concerning electrodynamic phenomena, meson-nucleon interactions and the physics of neutrinos**;

— and lastly, for the characteristic constants of nuclides: **questions concerning atomic weights, nuclear processes, etc.**

Of course, the Congress will be divided into several Sections, which will be specified later.

At the same time the XLII Congress of the Italian Physical Society will take place in Turin in its customary form.

Secretary General of the whole organization is prof. R. DEAGLIO, Institute of Physics of the University of Turin (via P. Giuria 1, Turin), to whom those wishing to attend the Congress may send their application or apply for information not later than April 30th 1956.

of EINSTEIN which lasted for ten years and finally came to a victorious end in 1915. The hystory of this period is an unprecedented chapter in the eternal struggle of the human intellect toward light and freedom.

Two fortunate circumstances came together. The one was that EINSTEIN had an intimate friend in the person of the Swiss mathematician MARCEL GROSSMANN, a mind of quick and flexible perception whom EINSTEIN could consult in all questions of mathematical technique. The other was that EINSTEIN himself was no matematician. Thus the danger of trivializing the problem by accepting a purely formal solution—any mathematical statement can be formulated in any arbitrary reference system—was avoided.

Already early in the game the phenomenon of *gravity* came in the focus of EINSTEIN's interest. It was natural that going beyond the group of reference systems of uniform motion he first chose as simplest example the reference systems which were in uniform *acceleration*. And now he noticed that the introduction of such a reference system generated an apparent field of gravity. EINSTEIN's famous « equivalence principle » [9] expressed the following thought. A physicist makes measurements in a closed laboratory where a uniform field of gravity exists, due to the gravitational attraction of the earth. Suddenly the earth is blasted away and the field of gravity destroyed. At this moment a giant begins to pull the entire laboratory upwards with the uniform acceleration g. Then the physicist will not notice anything of the outside happenings. Everything will occur in his laboratory exactly as it occurred before the blast. This thought experiment gave an important clue in EINSTEIN's hand. In the apparent gravitational field generated by the pull of the giant it is entirely natural that all bodies fall toward the earth with the same acceleration and that the force which they exert on the table on account of their weight, is strictly proportional to their inertial mass. Hence the astonishing fact, demonstrated to a very high degree of experimental evidence by the Hungarian physicist EÖTVÖS' gravimetrical measurements, that heavy mass and inertial mass are strictly proportional, became a natural consequence of the equivalence hypothesis.

EINSTEIN noticed that the Minkowskian line-element which expressed the uniform propagation of electromagnetic waves with the velocity c in every direction, changed its form considerably if an accelerated system was introduced. But even in the most arbitrary systems, which introduced motions of a most outlandish character, provided that the principle of continuity was preserved, the line-element still maintained a very definite character; it always remained a quadratic differential form of the variables, although the coefficients were no longer constants but became some more or less complicated functions of the coordinates:

$$ds^2 = g_{ik}\,dx_i\,dx_k.$$

If we write a formula in this form, we make use of « Einstein's sum convention » which helped so tremendously in taking out the drudgery from the formulae of absolute calculus. The sum convention is that we should automatically sum over two equal indices, without indicating the summation specifically by a Σ-sign.

Now EINSTEIN started to plunge wholeheartedly into the intricacies of absolute calculus which is the mathematical tool for the investigation of arbitrary reference systems [12]. This branch of mathematical analysis takes its departure point from the ingenious geometrical discoveries of GAUSS and RIEMANN and found further nourishment in the studies of CHRISTOFFEL, LIPSCHITZ, BELTRAMI and others. By a fortunate circumstance an elegant codification of the entire edifice of « tensor calculus » was given a few years earlier (in 1900) by the Italian mathematicians RICCI and LEVI-CIVITA (5). Probing into the possibilities of this calculus he became more and more lost in a strange land which for the first did not show any similarity to the customary constructions of theoretical physics. But he did not lose courage since the equivalence hypothesis showed him clearly that the mystery of universal gravity must have something to do with those g_{ik} quantities which form the components of the metrical tensor. The metrical tensor! How peculiar that all equations of special relativity, if they were written down in arbitrary coordinate systems, displayed prominently the metrical tensor. It looked as if this metrical tensor were part of the problem, as if it belonged to the unknowns of the problem. But then, since the g_{ik} are continuous field functions of the coordinates, we should expect field equations for their determination. How shall we find these equations? They certainly must have « general covariance », i.e. they must be expressible in any arbitrary frame of reference. Hence they must be constructed according to the rules of absolute calculus. The simplest covariant differential operator was of the second order and led to the fundamental curvature tensor of Riemann, R^i_{jkm}. If this tensor was put equal to zero, one obtained the flat Minkowskian space and that was certainly not desirable since then all gravitational fields would be reducible to mere apparent fields and that is certainly impossible. Now the Riemann tensor can actually not be expected to be suitable for the establishment of field equations since it overdetermines the problem. It has too many components. In four dimensions we have 10 independent g_{ik} and 20 independent R^i_{jkm}. We cannot expect 20 equations for 10 unknowns.

EINSTEIN overcame the difficulty by an ingenious device. A simple contraction reduced the order of the tensor from four to two. He now obtained the tensor

$$R_{ik} = R^{\alpha}_{i\,\alpha k}\,.$$

(5) *Math. Annalen*, **54**, 125 (1900).

This tensor had exactly the right degree of determination. It was neither over-determined nor under-determined. It was a symmetric tensor of second order, just as the tensor $g_{ik} = g_{ki}$ itself. We will thus get 10 equations for 10 unknowns. And even more miraculously: a slight modification of this tensor led to a new tensor which had a further fundamental property. The tensor

$$S_{ik} = R_{ik} - \tfrac{1}{2} R g_{ik} ,$$

was not only a symmetric tensor of second order but *its divergence was automatically zero*, expressing a conservation law of the same kind as the conservation law of momentum and energy in field-theoretical formulation. Hence a tensor was constructed in a very natural way out of the fundamental elements of Riemannian geometry which had exactly the same properties as the « matter tensor » of physics, encountered earlier in MINKOWSKI's work. EINSTEIN could thus establish the relation

$$S_{ik} = \varkappa T_{ik} ,$$

where $\varkappa$ is a constant whose value depends on the units in which lengths and weights are measured. Its relation to Newton's gravitational constant k came out as follows:

$$\varkappa = \frac{8\pi k}{c^4} .$$

This equation represents a triumph of speculative-constructive thinking which is unparalleled in the entire history of science. If special relativity (Einstein's theory of 1905, so called in distinction to his theory of 1915) united space and time into one inseparable entity, general relativity now added physical matter to the realm of geometry by showig how matter can be conceived in terms of curvature. The fundamental equation of Newtonian potential

$$\Delta \varphi = - 4\pi \varrho ,$$

which connected the gravitational potential with the density of matter, was enlarged to a much more comprehensive scheme of ten equations in which the scalar density of matter was replaced by the ten components of the matter tensor and the scalar Newtonian potential by the ten components of the metrical tensor. And yet, when EINSTEIN investigated these equations under the simplified conditions which hold if matter is very weak and distributed in spherically symmetric fashion, he found that the theory leads to exactly

the same conclusions as the much simpler scalar Newtonian theory. The entire Newtonian theory, based on the two pillars:

1) Mass times acceleration equal moving force.

2) The moving force between two masses is proportional to the product of the two masses and inversely proportional to the square of their mutual distance,

was now deducible from mere speculation, including the strict proportionality of inertial and gravitating mass which was never explained before [16].

The uncanny feature of the new theory was that while before science was happy if it succeeded in finding *some* mathematical law which could describe a certain body of observations, now the mathematical law itself became deducible from purely logical operations. Man in his eternal up-hill fight for knowledge has reached a new plateau.

We have telescoped a dramatic struggle of ten years into the scope of a few sentences. This is made possible by the fact that today the entire body of knowledge which is needed for Einstein's theory, is common property of every graduate student of physics. To obtain the proper perspective for the immense achievement we should visualize that at the time when EINSTEIN started on his journey, the juggling with curvilinear coordinates—a complete routine matter today—was hardly known, the absolute calculus was not studied and Riemannian geometry was not much more than a mathematical dream. To model this untractable material into a shape in which the outlines of a new cosmic order dawned on us and finally a whole symphony of new ideas dazzled the eye, required more than human efforts. At the end of his essay: « Origins of the general theory of relativity » [45] which gives a brief account of his struggles, EINSTEIN finishes the narrative with the following words: « In the light of finally obtained knowledge the deductions seem almost self-evident and can be understood with no great difficulties by any intelligent student. But the foreboding search in the dark, year after year, with its intense yearnings, its alternation from confidence to despondence, and then the ultimate break-through to final clarity, can only be perceived by somebody who experiences it himself ».

More than once he was lured into impassable cul-de-sacs. Once he was very near to the final solution and only a small step separated him from victory [13]. But when he tried to solve the field equations under the simplified conditions which hold for weak matter [17], he was discouraged by a peculiar phenomenon. He expected from the experiences of other field theories that the field equations will determine the g_{ik} uniquely if the proper initial and boundary conditions are prescribed. But actually an infinity of solutions existed, in contradiction to the causality principle of nature. This puzzle

caused a relapse of more than two years. In fact he thought that from this phenomenon the conclusion may be drawn that the field equations of gravity cannot be generally covariant [13], i.e. the principle of general relativity has to be sacrificed in favor of some restricted class of reference systems. In actual fact it is entirely natural that in view of the arbitrariness of reference systems the solution cannot be unique as long as we do not specify the reference system we are going to use. A contradiction to causality would only come about if more than one solution would exist in *one and the same* reference system. This, however, is not the case. « How simple », we may be inclined to say. But how absolutely wrong we are if we try to set up rational standards in an unexplored country in which every tree, every blade of gras sassumes a new shape and color and every foreign noise frightens us into panic, although it may only be the murmur of a creek.

After the final break-through EINSTEIN experienced an additional happy surprise which filled him with almost ecstatic joy. When he calculated the motion of a planet around the sun in second approximation, he found a slight deviation from the Newtonian theory in the form of a precession of the perihelion, given by the following expression [14]:

$$\eta = 6\pi \frac{kM}{c^2 a(1-e^2)},$$

(M = mass of sun, a = half-major axis, e = excentricity). This amount evaluated for the planet Mercury, gave exactly that mysterious surplus-precession of 43″ per century which remained as an unexplicable rest-effect, after taking into account all the perturbations (LEVERRIER, 1856). A few years ago G. M. CLEMENCE of the U.S. Naval Observatory subjected the entire perihelion problem to a critical re-examination. His calculations led to the following result, quoted from the author's summary (6): « The theoretical relativity effect of the motion of Mercury's perihelion is 43.03″ ±0.03″; the value obtained by subtracting all other known effects from the total observed motion is 42.56″ ±0.94. For the earth's perihelion the corresponding figures are 3.8″ ±0.0″ and 4.6″ ±2.7″. The confirmation by observation of the relativity effects is regarded satisfactory for both Mercury and the earth ».

Apart from the precession of the perihelion, Einstein's theory predicted two further optical effects. The one was a red-shift of the spectral lines if the emitting atom is on the surface of a star of the radius r_0 and the mass M:

$$\frac{\Delta\lambda}{\lambda} = \frac{k}{c^2}\frac{M}{r_0}.$$

(6) *Rev. Mod. Phys.*, **19**, 361 (1947).

This red-shift is expressible in the form of v/c where v is an equivalent receding velocity which would cause (due to Doppler effect) the same shift of the spectral lines as the Einstein gravitational effect. This velocity is on the sun only 0.64 km/s. Hence it is difficult to separate this very small effect from other Doppler effects, caused by temperature and pressure. It is imperative that spectral lines shall be chosen which are singularly sharp and to a large degree insensitive to temperature and pressure effects. The originally negative results of Ch. E. St. John at the Mt. Wilson Observatory were caused by an improper selection of lines. A careful re-examination of the problem gave a complete confirmation of the Einsteinian prediction (7). Even more decisive was the spectacular result obtained by W. S. Adams of the Mt. Wilson Observatory on the famous « dark companion » of Sirius (8). Since the distance of Sirius is well known, we possess from the luminosity of the star its radius, but we also know its mass from its influence on the orbit of Sirius. Hence we can calculate the red-shift of the light coming from the dark companion. Since the mass is comparable to that of the sun, while the radius is comparable to that of the earth, the red-shift becomes magnified by the factor 40, in comparison to the red-shift on the sun. The measurements of Adams gave 23 km/s as the mean value of the gravitational Doppler effect, compared with the theoretically predicted 26 km/s.

The second optical effect predicted by Einstein's theory is a bending of lightrays, by the sun. The effect is twice as big as the bending of a particle by the sun, moving with light velocity and calculated by ordinary mechanics. The deviation for a light ray which passes the limb of the sun becomes

$$\delta = \frac{4k}{c^2}\frac{M}{r_0} = 0.847 \cdot 10^{-5} \quad = 1''.75 .$$

This amount, observable only at the time of a solar eclipse, is very small and taxes the accuracy of the measuring instruments to the utmost. A number of astronomical expeditions were organized to test this consequence of Einstein's theory. Two British expeditions to Sobral in 1919 found that the statistical reduction of their plates gave for the deflection constant $1.98'' \pm 0.12''$, respectively $1.61'' \pm 0.30$. The 1922 expedition of the Lick Observatory gave 1.72 ± 0.15. However, the expedition of the Einstein Tower Observatory of Potsdam, Germany, gave in 1929 the value 2.24 ± 0.10 (9) although Trumpler reduced the same observations by a different statistical method and obtained

(7) *Astrophys. Journ.*, **67**, 195 (1928). For a rejective discussion of the Einstein effects cf. E. F. Freundlich: *Wissensch. Ann.*, **1**, 35 (1952).

(8) *Proc. of the Nat. Acad. of Science*, **11**, 382 (1925).

(9) *Zeits. f. Astrophysik*, **3**, 171 (1931).

1.75 ± 0.13 ([10]). The Yerkes Observatory obtained in 1947 the value 2.01 ± 0.27, in 1952 the value 1.70 ± 0.10 ([11]). The last result is not considered reliable, however, because of poor atmospheric conditions.

Our solar system was not the only beneficiary of EINSTEIN's revolutionary ideas. The new insights into the geometrical structure of the space-time world had radical cosmic consequences as well. If one imagined that the masses concentrated in the stars created a certain average mass density in the universe, one could see that an average curvature had to be present in the world. This average curvature had the consequence that space bent back on itself and came to a close in a similar way as a circle comes to a close. Space was no longer infinite; it was a boundless but *finite* space in which a traveller eventually returned to his starting point if he continued his journey constantly in the same direction. The dimensions of space were huge but not infinite. This led to the resolution of certain grave difficulties which are inherent in the assumption of an open space. In an open space we are unable to distribute any finite amount of matter without setting up a center of the world. Moreover, the stars would be doomed to gradually dissipate into infinity, with no chance of an equilibrium. In a finite universe these difficulties disappear. EINSTEIN thought that it should be possible to construct a finite and balanced universe in which the stars remain in the average in relative rest to each other. For this purpose he enlarged his original field equations by the so-called « cosmological term » which for ordinary purposes is excessively small but which becomes of decisive importance if it comes to spaces of cosmic dimensions. The connection between curvature tensor and matter tensor now became

$$S_{ik} + \lambda g_{ik} = \varkappa T_{ik} ,$$

where λ is the so-called « cosmological constant » [21]. EINSTEIN assumed a pre-established harmony between the average density and the cosmological constant, in order to make a static universe possible. It was shown, however, by G. LEMAITRE ([12]) that the Einsteinian universe is not stable since it becomes more and more time-dependent under the influence of the slightest disturbance. Moreover, the observations of E. P. HUBBLE of the distant nebulae showed conclusively that the universe is not static but in a state of constant expansion. This opened new avenues of cosmological investigations, with contributions from DE SITTER, FRIEDMAN, LEMAITRE, ROBERTSON, MILNE and others. EINSTEIN himself was not inclined to cosmological speculations of any specific kind. He considered our present knowledge of the role of matter

([10]) *Science*, **75**, 538 (1932).
([11]) G. J. WHITROW: *The Observatory*, **75**, 166 (1955).
([12]) *Ann. Soc. Sci. Bruxelles*, **47**, 49 (1927).

in general relativity so vague that any extrapolation into billions of years in the past or in the future seemed to him of not more than mythological value. He lost even confidence in the cosmological constant when Hubble's observations disproved the hypothesis of a static universe [28]. But the conclusion that the physical space has an average curvature and is a closed and finite space, remains one of the most cherished cosmological consequences of Einstein's theory.

4. – The years between 1905 and 1925 were the heroic years in EINSTEIN's scientific life. We could call them the King-Midas-years since everything he touched turned to gold in his hands. The achievements were on a magnificent scale and the results so overwhelmingly convincing, in spite of their revolutionary novelty, that they were immediately codified in the annals of science as treasures which will never be abandoned again. We now come to the second half of EINSTEIN's scientific life. He still had thirty years in front of him. What will happen to a genius who has transcended so often the usual measure of inventiveness granted to a single individual?

The fires were still burning and the fervor and devotion with which he probed into the mysteries of the universe, remained unabated, until the last day of his life. But the great disruption of European science on account of the Nazi hordes who swept over country after country, burning and destroying the intellectual values of Western civilization, had its effect on EINSTEIN. Generously the United States gave him sanctuary and he always gratefully acknowledged the splendid work conditions under which he could continue his scientific activities. But the loss of his friends and of the incomparable mental stimulus which the city of Berlin offered in the two decades between 1910 and 1930, had something to do with his gradual dropping out of the limelights of contemporary physics. The friends and admirers of EINSTEIN often regretted the « splendid isolation » which became more and more his habit as the years went by. He said himself [43, p. 5]: « I live in that solitude which is painful in youth, but delicious in the years of maturity ».

The most important springs of this isolation, however, lay in himself and in the inevitable destiny which drowe him forward. The ten years in which he worked on general relativity, had a profound influence on his philosophy and changed substantially his scientific outlook. EINSTEIN started his scientific career as a cool and sceptical positivist, strongly influenced by the teachings of the Viennese philosopher MACH [46]. He put physical intuition highly above mathematical speculation and was inclined to see in mathematics nothing more than a purely formal tool of description which came about in MACH's sense as an accident of historical evolution rather than as a necessary part of nature itself. But if one looked at the final draft of the new gravitational theory, one could not help seeing that it was the result of philosophical

and mathematical speculations of the highest order and not the outcome of a mere « description of observations » to which the positivistic philosophy relegates the role of theoretical physics. No amount of experimentation could ever suggest a set of equations such as the Einsteinian field equations of gravity. No amount of descriptive activity could ever bring a physicist to the strange wonders of Riemann's curvature tensor. Under the impact of general relativity EINSTEIN abandoned the tenets of positivism and reverted to the old Greek natural philosophers who believed that nature is reasonable and explorable by mere philosophizing. Of course, the new mode of philosophizing appeared under much more exacting conditions. Geometry had no longer the simple pattern that it had in EUCLID's time. The most advanced mathematical tools had to be employed in order to penetrate into the true core of the universe. Thus EINSTEIN abandoned his earlier mistrust of mathematics and became a convert of abstract mathematical thinking—sometimes to a dangerous degree. But if we may regret that the ostracized himself to strange countries to which nobody was willing to follow him, we have to admire the intellectual honesty with which he pursued his aims. Einstein-Faust gazed at the sign of the Macrocosmos for ten years, trying magic formula after magic formula, to conjure up the mighty spirit of creation. And then the magic formula was found and the mighty spirit did appear in all its splendor. Can we blame him if he could not settle down again to the gray daily chores of life? Can we expect that a man who has seen face to face the cosmic apparition, should help little people in building little homes?

The challenge was too great to be satisfied with mere patchwork. Nature is certainly not composed of watertight compartments. If we were able to unriddle the mystery of gravity by inspired speculation, could not the same be done with the other forces of nature? Electricity is the most fundamental phenomenon of nature and EINSTEIN hoped that the structure of elementary particles could perhaps be explained as an equilibrium between the repulsive electric and the attractive gravitational forces. But what is electricity?

The tying up of the curvature tensor with a matter tensor of external origin was not to EINSTEIN's liking. All the great victories of his gravitational theory were made possible by the fact that matter is concentrated in very small portions of space. Hence it was essentially the equation

$$S_{ik} = 0 ,$$

which was needed in explaining the gravitational phenomena, without any specific « matter tensor ». This matter tensor was not more than a feeble substitute for that great unknown that lurks in the atomistic structure of material particles and that escapes our knowledge completely. If we would understand the essence of electricity in the same way as we understand the

essence of gravity, all difficulties would disappear and the mystery of the atom would be explained. Gravity alone is certainly not enough for the construction of a reasonable universe because the solution of the equations of gravity leads to infinities which cannot be tolerated in a logical description of nature [35]. Something corresponding to electricity is demanded to make the construction of the elementary particle possible. But what is the geometrical meaning of electricity?

The electromagnetic phenomena are described with the help of the Maxwellian equations and these equations are based on an anti-symmetric tensor of second order $F_{ik} = -F_{ki}$, in contrast to the metrical tensor which is symmetric: $g_{ik} = g_{ki}$. Hence EINSTEIN started to search for an anti-symmetric tensor of second order in Riemann's geometry which, however, he could not find. Well, then let us abandon Riemann's geometry in favor of a still more general geometry. But here the situation was very different from his earlier journey when he abandoned Euclidean geometry in favor of Riemannian geometry. In the earlier journey he was guided by his unfailing physical intuition. Moreover, the path was laid out mathematically by the magnificent inner consistency and self-evidency of tensor calculus which was so eminently well adapted to the formulation of physical laws. And thus the earlier journey was a constant climbing to higher and higher vistas and the reaching of higher and higher plateaus. The foundation was so solid that the danger of crashes was eliminated, even if one deviated sometimes from the straight road. EINSTEIN's later journeys are of a more jagged nature and marred by many reverses. It was now a groping in the dark and his physical intuition was of little help if the mathematical intricacies of the problem threatened to drown out the basic line of thought. The difficulty was that Riemannian geometry was logically the only possible heir to Euclidean geometry but if Riemannian geometry had to be abandoned, a number of throne-pretenders appeared on the scene.

The first theory which attempted at a geometrical unification of electricity and gravity was developed by H. WEYL ([13]). He based his deductions on the assumption that the real physical quantities are the *ratios* of the g_{ik} and not the g_{ik} themselves. This leaves a scalar free in the determination of the metric which can be represented with the help of a linear differential form $\varphi_\alpha dx_\alpha$. The φ_i are then interpreted as components of the electromagnetic vector potential. Weyl's geometry was only the first step toward a still wider group of geometries in which the line-element is altogether abandoned in favor of the Γ^m_{ik}-quantities which now become the fundamental elements of a non-metrical geometry in which the parallel displacement of vectors and tensors is still possible. These so-called « affine geometries » were extensively inves-

([13]) *Ann. der. Phys.*, **59**, 101 (1919).

tigated by EDDINGTON, EINSTEIN, and many others ([11]). In recent years Schrödinger made further advances and brought his researches in relation to the last unified field theory of EINSTEIN ([15]). EINSTEIN could never warm up to the idea of abandoning the line-element since he considered the very high persistency of spectral frequencies throughout the universe as a direct proof for the existence of an infinitesimal yardstick which can be transported without change from one point to the other. In the affine geometries there is no reason for the existence of such a rigid yardstick.

EINSTEIN's own efforts to combine electricity and gravity into one unified geometriscal scheme can be catalogued under three major headings.

Distant parallelism. Riemannian geometry does not allow the notion of a parallelism between distant points which is independent of the path. By introducing n infinitesimal and mutually orthogonal unit vectors $h_{\mu a}$ at every point of the manifold, we obtain 16 fundamental quantities from which the 10 g_{ik} can be constructed with the usual Riemann parallelism. We have, however, absolute parallelism too and this gives rise to a new curvature tensor [26]

$$\tfrac{1}{2}(\Delta^m_{ik} - \Delta^m_{ki}) = \Delta^m_{ik},$$

which is anti-symmetric in i, k. In terms of these quantities a formal interpretation of the Maxwell equations becomes possible. EINSTEIN lost confidence in the theory when the field equations, derived from the simplest Hamiltonian, allowed static solutions which corresponded to the fields of electrically charged particles, in contradiction to the Lorentz force which must bring such particles into motion [27]. Field equations without a Hamiltonian, on the other hand, remained too vague, on account of the many choices which are left open to them [29].

The five-dimensional world. TH. KALUZA conceived the idea of generalizing Einstein's equations from four to five dimensions and considering the motion of a particle, in analogy to Einstein's postulate in ordinary gravity, as a shortest line (geodesic) of this five-dimensional manifold ([16]). The fifth dimension entered in the line-element but the fifth variable remained dormant since the g_{ik} were supposed to be independent of the surplus variable. The motion of an electrically charged particle in an electromagnetic field (Lorentz-force) could thus be explained. But no clues could be drawn concerning the structure of elementary particles since the field equations did not allow solutions which

([11]) Cf. Sir E. WHITTAKER: *History of the Theories of Aether and Electricity, 1900-1926* (London, 1953), p. 190, for a condensed bibliography.

([15]) *Proc. Roy. Irish Acad.*, **49**, 43 (1943), and subsequent papers.

([16]) *Preuss. Akad. Wiss. Berlin*, Ber. 966 (1921).

avoided infinities [22]. VEBLEN and HOFFMANN gave in 1930 a different interpretation of the fifth dimension, by changing from a metrical to a projective geometry ([17]). EINSTEIN arrived independently somewhat later to an equivalent theory [30, 31], by introducing a flat five-dimensional space in addition to the curved four-dimensional space, and defining the relations which shall exist between the parallel displacements in these two types of spaces. He thus obtained a formal interpretation of the electromagnetic energy tensor. In further pursuance of these ideas he once more returned to Kaluza's original theory [33] but replaced the unnatural assumption that the fifth variable remains completely inactive in the structure of the world, by a more natural hypothesis.

The non-symmetrical line-element. In the last years of his life EINSTEIN returned to the four-dimensional structure of the world but thought to enlarge the ordinary point-transformations of general relativity by transformations involving complex coefficients. If we do so, the g_{ik} cease to be real functions of the variables but become complex numbers, yet in such a way that the entire line element remains real; (Hermitian forms). The symmetry condition $g_{ik} = g_{ki}$ is now replaced by the Hermitian condition [36]

$$g_{ik} = \overset{*}{g}_{ki} .$$

The real part of the g_{ik} give a symmetric, the imaginary part an antisymmetric tensor. They can be correlated to gravitation and electromagnetism respectively. Later EINSTEIN abandoned the complex nature of the line-element and assumed a real but non-symmetric g_{ik}. If the g_{ik} became symmetric, one returned to the original gravitational equations. The anti-symmetric part of the g_{ik}, however, could be made responsible for the appearance of electricity in nature. This theory went through several modifications but the basic ideas remained unchanged [37, 40]. They represent the last efforts of EINSTEIN to penetrate to the ultimate secrets of the universe. His last paper on the unified field theory came out shortly after his death [41].

For the final evaluation of EINSTEIN the scientist it makes not the slightest difference whether he did or did not find the ultimate equations of the universe. He discovered such a galaxy of dazzling ideas and gave mankind a program of such nobility that his unique position in the history of human thought is ensured for all ages.

5. – We will conclude this sketch of the scientific evolution of ALBERT EINSTEIN with a brief summary of his life-data [46, 47]. He was born in Ulm (Germany), on March 14, 1879, but his parents soon moved to Munich where

([17]) *Phys. Rev.*, **36**, 810 (1930).

he received his primary and secondary schooling. He did not like school, nor did school like him; (his Latin professor told him once: « Einstein, you will never amount to anything »). His great liking for the natural sciences and for geometry was clearly marked from his earliest youth. When his father failed in his engineering business and emigrated to Italy, EINSTEIN went to Switzerland in order to finish his secondary education. The Swiss form of democracy and its freedom from any military pressure was much more to his liking than the nationalistic and militaristic tradition of Germany. In 1901 he acquired Swiss citizenship. His college years were spent at the Polytecnicum of Zurich where he graduated in 1901. In 1902 he became a Swiss civil servant by accepting a minor position in the Patent Office of Berne. Here he remained up to 1909. His duties were in no way inspiring but left him much time for day-dreaming in physics. He got married in 1902 and two sons were born to him. However, the marriage was not a success and was dissolved in 1919. In 1909 he was already an internationally known physicist of the first rank. He received a call to the University of Zurich, where he remained up to 1914, with an interlude of one year in 1911-12, spent as full-professor of physics at the University of Prag (Austria). In 1914, at the instigation of MAX PLANCK and supported by NERNST, WARBURG and RUBENS, EINSTEIN was elected a member of the Berlin Academy of Sciences, under extraordinarily favorable conditions. He was made Director of the Research Institute for Physics of the Kaiser-Wilhelm-Institute in Berlin. Free of any official obligations, free even of any lecturing duties, he could devote himself fully to the pursuance of his speculations. His years in Berlin (from 1914 to 1933), were extraordinarily happy. He attended regularly the physics seminars of the university. These seminars saw a galaxy of stars of first magnitude which filled the audience with awe: EINSTEIN, PLANCK, LAUE, NERNST, SCHRÖDINGER, RUBENS were almost always present. The lively discussions which developed after every presentation—reviews of the latest stand of experimental and theoretical research in all branches of physics—remain in the indelible memory of everyone who attended this incomparable gathering. In 1919 EINSTEIN married his cousin Elsa with whom he lived happily until her death in 1936. When the great Dutch physicist H. A. LORENTZ retired in 1923 from his chair of theoretical physics at the University of Leyden, EINSTEIN was appointed as his successor. Since, however, he did not want to lose his Berlin associations, he consented to go to Leyden every year for a temporary visit, staying with his intimate friend EHRENFEST and giving lectures at the university as a visiting professor. EINSTEIN was fortunately abroad when in 1933 the Nazis came to power. Recognizing the barbarous character of the new regime, he immediately resigned from his Academy post and renounced his German citizenship. The State confiscated his possessions. Already in 1932 the newly founded Princeton Institute for Advanced Study elected him to his faculty. This position

became now permanent. EINSTEIN emigrated to the United States and became U.S. citizen in 1940. In 1945 he reached retiring age but was requested to stay. He, however, did not want that an exception should be made in his favor and went into retirement, although continuing with his daily studies. His Princeton years were principally devoted to the motion problem of general relativity [32, 34] and to an extension of his gravitational theory to the inclusion of electricity. His efforts culminated in the « unified field theory » which went through several formulations. The last formulation which he considered the final one, came out after his death [41]. His health deteriorated since the end of 1954, but he continued his regular routine, although now working mostly at home.

The shadows were lengthening as the years passed by. The raging intellectual fire burned out his organism prematurely. EINSTEIN looked always about ten years older than he actually was. More than once in his life he was dangerously ill but always with a good chance of recovery. During the last year, however, the rapid decline of his physical forces became alarmingly manifest. When in April of this year he was transferred to the hospital of Princeton, he knew that no hope was left. Calmly and unperturbed, with no pathos, no sentimentality, no regret, he waited for the approaching death. He died with the same simplicity and humility with which he lived. « Homo liber de nulla re minus quam de morte cogitat », said the great Dutch philosopher SPINOZA whom EINSTEIN held in particularly high esteem. « The free man thinks of nothing less than of death ». ALBERT EINSTEIN was a free man!

BIBLIOGRAPHY

An excellent list of EINSTEIN's scientific writing (up to 1949, 309 items), can be found in [44], p. 694-730. The following selective bibliography is only a cross-sectional enumeration of the most important papers, chosen in view of the references of the article. The abbreviation *Berl. Ber.* refers to: *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Berlin.*

[1] *Kinetische Theorie des Wärmegleichgewichtes und des zweiten Hauptsatzes der Thermodynamik*, in *Ann. der Phys.*, **9**, 417 (1902).

[2] *Theorie der Grundlagen der Thermodynamik*, in *Ann. der Phys.*, **11**, 170 (1903).

[3] *Allgemeine molekulare Theorie der Wärme*, in *Ann. der Phys.*, **14**, 354 (1904).

[4] *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*, in *Ann. der Phys.*, **17**, 132 (1905).

[5] *Die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, in *Ann. der Phys.*, **17**, 549 (1905).

[6] *Elektrodynamik bewegter Körper*, in *Ann. der Phys.*, **17**, 891 (1905).

[7] *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?*, in *Ann. der Phys.*, **18**, 639 (1905).

[8] *Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme*, in *Ann. der Phys.*, **22**, 190 (1907).

[9] *Relativitätsprinzip und die aus demselben gezogenen Forderungen*, in *Jahrbuch der Radioaktivität*, **4**, 411 (1907).
[10] *Zum gegenwärtigen Stande des Strahlungsproblems*, in *Phys. Zeits.*, **10**, 185 (1909).
[11] *Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung*, in *Phys. Zeits.*, **10**, 817 (1909).
[12] *Entwurf einer verallgemeinerten Relativitätstheorie und eine Theorie der Gravitation.* I. *Physikalischer Teil* von A. EINSTEIN; II. *Mathematischer Teil* von M. GROSSMANN, in *Zeits. f. Math. u. Phys.*, **62**, 225 (1913).
[13] *Kovarianzeigenschaften der Feldgleichungen der auf die verallgemeinerte Relativitätstheorie gegründeten Gravitationstheorie*, in *Zeits. f. Math. u. Phys.*, **63**, 215 (1914).
[14] *Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie*, in *Berl. Ber.*, 831 (1915).
[15] *Feldgleichungen der Gravitation*, in *Berl. Ber.*, 844 (1915).
[16] *Grundlagen der allgemeinen Relativitätstheorie*, in *Ann. der Phys.*, **49**, 769 (1916).
[17] *Näherungsweise Integration der Feldgleichungen der Gravitation*, in *Berl. Ber.*, 688 (1916).
[18] *Hamiltonsches Prinzip und allgemeine Relativitätstheorie*, in *Berl. Ber.*, 1111 (1916).
[19] *Zum Quantensatz von Sommerfeld und Epstein*, in *Verhandlungen der Deutschen Physikalischen Gesellschaft*, **19**, 82 (1917).
[20] *Quantentheorie der Strahlung*, in *Phys. Zeits.*, **18**, 121 (1917).
[21] *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*, in *Berl. Ber.*, 142 (1917).
[22] *Beweis der Nichtexistenz eines überall regulären zentrisch symmetrischen Feldes nach der Feldtheorie von Kaluza*, (with J. GROMMER), in *Jerusalem University, Scripta*, **1**, No. 7, 5 pp.
[23] *Quantentheorie des einatomigen idealen Gases*, in *Berl. Ber.*, 261 (1924).
[24] *Quantentheorie des einatomigen idealen Gases*, 2. Abhandlung, in *Berl. Ber.*, 3 (1925).
[25] *Kaluzas Theorie des Zusammenhanges von Gravitation und Elektrizität*, in *Berl. Ber.*, 23 (1927).
[26] *Auf die Riemann-Metrik und den Fern-Parallelismus gegründete einheitliche Feld-Theorie*, in *Math. Ann.*, **102**, 685 (1930).
[27] *Zwei strenge statische Lösungen der Feldgleichungen der einheitlichen Feldtheorie*, (with W. MAYER), in *Berl. Ber.*, 110 (1930).
[28] *Zum kosmologischen Problem der allgemeinen Relativitätstheorie*, in *Berl. Ber.*, 235 (1931).
[29] *Systematische Untersuchung über kompatible Feldgleichungen welche in einem Riemannschen Raume mit Fern-Parallelismus gesetzt werden können*, (with W. MAYER), in *Berl. Ber.*, 257 (1931).
[30] *Einheitliche Theorie von Gravitation und Elektrizität*, (with W. MAYER), in *Berl. Ber.*, 541 (1931).
[31] *Einheitliche Theorie von Gravitation und Elektrizität*, 2. Abhandlung, (with W. MAYER), in *Berl. Ber.*, 130 (1932).
[32] *Gravitational equations and the problem of motion*, (with L. INFELD and B. HOFFMANN), in *Ann. of Math.*, **39**, 65 (1938).
[33] *Generalization of Kaluza's theory of electricity*, (with P. BERGMANN), in *Ann. of Math.*, **39**, 683 (1938).
[34] *Gravitational equations and the problem of motion*, II, (with L. INFELD), in *Ann. of Math.*, **41**, 455 (1940).
[35] *Demonstration of the non-existence of gravitational fields with a non-vanishing total mass free of singularities*, in *Rev. Univ. Tucuman*, A, **2**, 11 (1941).

[36] *Generalization of the relativistic theory of gravitation*, in *Ann. of Math.*, **46**, 578 (1945).
[37] *Generalized theory of gravitation*, in *Rev. Mod. Phys.*, **20**, 35 (1948).
[38] *The Bianchi identities in the generalized theory of gravitation*, in *Canad. Journ. of Math.*, **2** 120 (1950).
[39] *A comment on a criticism of unified field theory*, in *Phys. Rev.*, **89**, 321 (1953).
[40] *The meaning of relativity*; 4th ed. (with Appendix II) (Princeton, N. J., 1953).
[41] *A new form of the general relativistic field equations*, (with B. KAUFMAN), in *Ann. of Math.*, **62**, 128 (1955).

Articles and books of a non-technical nature.

[42] *The World As I See It* (New York, 1934); 290 pp.
[43] *Out of My Later Years* (New York, 1950); 279 pp.
[44] *Albert Einstein: Philosopher-Scientist* (Evanston, Ill., 1949); 781 pp. In this outstanding collection of essays about EINSTEIN the following articles have particular bearing on the contents of the present paper:
A. EINSTEIN: *Autobiographical Notes.*
L. DE BROGLIE: *The Scientific Work of Albert Einstein.*
W. PAULI: *Einstein's Contributions to Quantum Theory.*
M. BORN: *Einstein's Statistical Theories.*
M. VON LAUE: *Inertia and Energy.*
[45] *Mein Weltbild* (Zürich, 1953); 275 pp.
[46] P. FRANK: *Einstein, His Life and Times* (London, 1948); 367 pp.
[47] C. SEELIG: *Albert Einstein* (Zürich, 1954); 304 pp.

FONDAZIONE FRANCESCO SOMAINI
PRESSO IL TEMPIO VOLTIANO, A COMO

BANDI DI CONCORSI
AL PREMIO E ALLA BORSA PER IL 1958

Con lo scopo di premiare e incoraggiare nel nome di ALESSANDRO VOLTA gli studi di Fisica in Italia, la « Fondazione Francesco Somaini », presso il Tempio Voltiano a Como, indice i seguenti concorsi.

A) **Concorso al " Premio Triennale per la Fisica Francesco Somaini " per il 1958.** di L. 1 500 000 (un milione e cinquecentomila) nette, da assegnarsi al concorrente che, fra quelli che la Commissione Giudicatrice giudicherà in senso assoluto meritevoli del premio per i risultati conseguiti nello studio della Fisica durante il Triennio 1° Luglio 1955-30 Giugno 1958, sia, a parere della Commissione stessa, il più meritevole.

B) **Concorso alla " Borsa Francesco Somaini per lo studio della Fisica " per il 1958.** di L. 750 000 (settecentocinquantamila) nette, da assegnarsi al concorrente che, tra quelli che la Commissione Giudicatrice giudicherà in senso assoluto meritevoli della Borsa, verrà dalla Commissione stessa giudicato il più meritevole, sia per titoli, preparazione scientifica, lavori già svolti e risultati già conseguiti nella Fisica, sia anche per il vantaggio che gli studi, per i quali è richiesta la Borsa, possono portare allo sviluppo della Fisica, in Italia.

1. – Ad entrambi i Concorsi possono prendere parte singolarmente i cittadini, d'ambo i sessi, italiani e svizzeri del Canton Ticino purchè di stirpe italiana. Sono esclusi dal Concorso i membri della Commissione Amministratrice e della Commissione Scientifica della « Fondazione Francesco Somaini ».

2. – Le norme particolareggiate dei singoli Concorsi verranno pubblicate in apposito volantino che potrà essere richiesto dagli interessati alla Segreteria della Fondazione presso il Tempio Voltiano a Como.

3. – La domanda, i documenti, i lavori, ecc., presentati dai singoli concorrenti dovranno pervenire, tra il 1° Gennaio e le ore 12 del 1° Luglio 1958, alla Commissione Amministratrice della « Fondazione Francesco Somaini » a Como presso il Tempio Voltiano.

4. – Il Premio Triennale per la Fisica potrà essere anche conferito a uno studioso che non abbia preso parte al Concorso, ma sia stato segnalato da un Membro della Commissione Giudicatrice, con proposta motivata, come meritevole di particolare considerazione oppure ritenuto degno di premio dalla Commissione Giudicatrice, indipendentemente da ogni segnalazione.

* * *

La procedura dei suddetti Concorsi è regolata secondo lo Statuto della Fondazione, il quale è ostensibile a Como presso il Tempio Voltiano ed è depositato negli atti del Notaio Dr. Raoul Luzzani di Como.

Como, dal Tempio Voltiano, il giorno 6 Agosto 1955.

Il Segretario Conservatore del Tempio
CESARE MORLACCHI

Il Presidente Sindaco di Como
PAOLO PIADENI

[Il bando in esteso di detti Concorsi è stato pubblicato nel fascicolo di Gennaio del *Nuovo Cimento* 1956].

SOCIETÀ ITALIANA DI FISICA

CONCORSO SPECIALE A UN PREMIO PER STUDI SUI MATERIALI SEMICONDUTTORI

Prendendo occasione dalla prossima inaugurazione a Forlì del Centro di Ospitalità per scrittori, artisti e scienziati, fondato, in memoria di Livio e Maria Garzanti, dal figlio, l'Editore Aldo Garzanti, questi ha istituito un premio scientifico, del quale ha fissato la natura e lo scopo con la sequente dichiarazione:

« Considerando la grande importanza che in questi ultimi anni hanno acquistato, sia nel campo strettamente scientifico sia nelle applicazioni, i materiali semiconduttori, viene istituito un premio di un milione di lire a favore di uno studioso italiano che, entro il 31 Maggio 1956, abbia presentato una memoria, nella quale, premesso un breve riassunto sulla Fisica di questi materiali, siano indicate, nei termini più concreti, le possibilità di fabbricazione e di impiego di essi in Italia, con riferimento alle materie prime qui esistenti, ai costi e alle condizioni della nostra industria. Sarà tenuto conto dei contributi scientifici e tecnici personali che il candidato abbia eventualmente portato alla conoscenza delle proprietà e delle applicazioni di questi materiali. Il premio sarà assegnato alla migliore tra le memorie giudicate degne dalla apposita Commissione. La proprietà letteraria e il diritto di eventuale pubblicazione della memoria vincitrice spettano all'Editore Garzanti ».

La Società Italiana di Fisica, richiesta dall'Editore Garzanti, ha ben volentieri acconsentito, applaudendo alla iniziativa di lui, a patrocinare il Concorso e curare tutte le operazioni che esso comporta: e quindi d'intesa con l'Editore stesso ha formulato il seguente Bando.

1. – È aperto un concorso a un premio di L. 1000000 (un milione) intitolato « Premio Livio e Maria Garzanti », per studi sui materiali semiconduttori.

2. – Possono partecipare al Concorso le persone di nazionalità italiana che non facciano parte della Commissione esaminatrice.

3. – Le domande di ammissione al Concorso, redatte su carta libera e recanti le generalità e il recapito del concorrente, dovranno pervenire alla Presidenza della Società Italiana di Fisica (Milano, Via Saldini, 50) entro il 31 Maggio 1956.

Alla domanda il concorrente dovrà unire: *a*) la dichiarazione firmata che egli è di nazionalità italiana; *b*) cinque copie (dattilografate) di una memoria originale, da lui compilata, nella quale, premesso un breve riassunto sulla Fisica dei materiali semiconduttori, siano indicate, nei termini più concreti, le possibilità di fabbricazione e d'impiego di essi in Italia, con riferimento alle materie prime qui esistenti, ai costi e alle condizioni della nostra industria; *c*) quei titoli, documenti e pubblicazioni (queste in cinque copie) che possano attestare il contributo portato dal concorrente alla conoscenza delle proprietà fisiche e alle applicazioni tecniche e pratiche dei materiali semiconduttori; *d*) sei copie (dattiloscritte) dell'elenco completo di tutte le carte presentate.

4. – La Commissione giudicatrice del concorso è costituita da cinque membri, di cui quattro nominati dal Consiglio di Presidenza della Società Italiana di Fisica e uno dall'Editore Aldo Garzanti in sua rappresentanza.

5. – Il Premio sarà assegnato alla migliore tra le memorie giudicate degne dalla Commissione. La proprietà letteraria e il diritto dell'eventuale pubblicazione della memoria vincitrice spettano all'Editore Aldo Garzanti.

6. – Il giudizio della Commissione è inappellabile.

7. – La proclamazione del vincitore avrà luogo, per quanto in tempo, alla inaugurazione del Centro di Ospitalità di Forlì, altrimenti, insieme con il conferimento del Premio stesso al Congresso Internazionale di Fisica che si terrà nel prossimo Settembre a Torino per celebrare il grande fisico e chimico italiano Amedeo Avogadro nel centenario della sua morte.

Il Segretario: G. C. Dalla Noce — *Il Presidente*: G. Polvani

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 5, 1955
2° Semestre

Carl Friedrich Gauß (*).

M. KRAFFT

Philosophische Facultät der Philipps-Universität, Marburg an der Lahn - Deutschland

Jeder, der es unternimmt, über die wissenschaftlichen Leistungen von CARL FRIEDRICH GAUSS su schreiben, sieht sich in einer sehr unbehaglichen Lage und bereut sein Unterfangen je länger, je mehr. Dabei sind die Unterlagen für eine solche Arbeit vollständiger und umfangreicher als sonst in entsprechenden Fällen. Es steht nicht nur die große, auch den Nachlaß verarbeitende zwölfbändige Ausgabe der Werke [1] zur Verfügung, deren Vollendung und hohen Wert man wesentlich der Tatkraft von FELIX KLEIN verdankt, sondern die Bände X_2 und XI_2 der Werke bestehen ganz aus eingehenden Würdigungen der Leistungen von GAUSS auf seinen verschiedenen Arbeitsgebieten durch hervorragende Fachgelehrte [2]. Dazu kommen die umfangreichen Veröffentlichungen des Briefwechsels von GAUSS [3]. Eine sehr eindrucksvolle kurze zusammenfassende Würdigung der Leistungen von GAUSS verdankt man F. KLEIN [4]. Die erste große Schwierigkeit ist nun, daß sich die Arbeit von GAUSS auf so viele verschiedene Gebiete erstreckt, die heute wohl niemand mehr gleichmäßig beherrscht. Die zweite ist die Fülle des Stoffes, die jeder Darstellung durch die nötige Auswahl eine stark subjektive Färbung gibt. Erschwert wird die Auswahl noch durch die überraschenden Verbindungen scheinbar heterogener Fragen. Am unangenehmsten ist aber die dritte Schwierigkeit: GAUSS hatte leider die Gewohnheit, zahlreiche seiner bedeutendsten Untersuchungen nicht zu veröffentlichen. Er hat dafür manchen berechtigten Tadel von seinen Freunden hinnehmen müssen. So schrieb ihm 1833 der Astronom SCHUMACHER, auf die Devise des Gaußschen Siegels « Pauca sed matura » anspielend: « Ich komme wieder auf meinen alten Wunsch zurück, daß es Ihnen gefallen möge, länger und schneller zu schreiben. Ich würde Ihnen dann sogleich ein neues Petschaft zu überreichen wagen mit "multa nec im-

(*) Zum Gedächtnis anlässlich des hundertsten Todesjahres.

matura,, ». Alle diese zurückgehaltenen Ergebnisse wurden z.T. noch bei Lebzeiten von GAUSS, z.T. in den nächsten Jahrzehnten nach seinem Tode von anderen Forschern wieder gefunden. Dabei haben in einzelnen Fällen versteckte Hinweise von GAUSS eine wichtige Rolle gespielt. Die lebendige Entwicklung knüpft überall an diese Neuentdecker an. Für die Kenntnis der Persönlichkeit von GAUSS ist es unerläßlich, auch auf diese von ihm nicht veröffentlichten Arbeiten einzugehen, man muß aber vermeiden, den späteren Forschern ihr Verdienst zu schmälern.

* * *

CARL FRIEDRICH GAUSS wurde am 30. April 1777 in Braunschweig als Sohn armer Eltern geboren. Schon in der Grundschule fiel er seinen Lehrern durch seine Begabung auf. Sie rangen daher dem widerstrebenden Vater die Zustimmung zu dem durch Stipendien ermöglichten Übergang seines Sohnes auf das Gymnasium ab. Auch die Lehrer des Gymnasiums und des weiterführenden Carolinums (der Keimzelle der heutigen Technischen Hochschule) nahmen sich des jungen GAUSS vorbildlich an. Vor allem empfahlen sie ihn dem regierenden Herzog KARL WILHELM FERDINAND. Dieser gewährte GAUSS eine für das arme Land sehr großzügige finanzielle Unterstützung, die im Laufe der Zeit auf die für die damaligen Verhältnisse erhebliche Summe von 600 Talern jährlich anstieg und an keine Verpflichtung irgend welcher Art gebunden war. Erst mit dem Zusammenbruch der alten Staatsordnung durch den Sieg NAPOLEONS über Preußen bei Jena und Auerstedt 1806 und dem Tode des Herzogs an der in der Schlacht erlittenen Verwundung nahm diese landesväterliche Fürsorge ein Ende. 1795-1798 besuchte GAUSS die Universität Göttingen, durch deren Bibliothek er EULER, LAGRANGE und die anderen großen Forscher kennenlernte, kehrte dann nach Braunschweig zurück und promovierte 1799 bei J. F. PFAFF an der braunschweigischen Landesuniversität Helmstedt. 1807 wurde er noch vor Gründung des kurzlebigen napoleonischen Königreichs Westfalen als Professor der Astronomie und Direktor der Sternwarte nach Göttingen berufen. In dieser Stellung blieb er bis zu seinem Tod am 23. Februar 1855. Die ersten Jahre in Göttingen waren durch Not infolge von Kriegskontributionen und durch das Ausbleiben oder die verspätete Zahlung des Gehaltes, verursacht durch die Verschwendung des Hofes und die ununterbrochenen Kriege, getrübt. Dazu kam viel familiäres Unglück, vor allem 1809 der Tod seiner heiß geliebten ersten Frau und 1831 der seiner zweiten Gattin.

* * *

GAUSS ist in der Mathematik und Astronomie und den andern von ihm gepflegten Wissenschaften wesentlich Autodidakt, genau wie die ganze ältere Generation seiner wissenschaftlichen Freunde und Korrespondenten, die alle

aus anderen Berufen herkamen. Es genügt, an die Astronomen Olbers, v. Zach, Lindenau, Schumacher, Bessel und den Mathematiker Pfaff zu erinnern. Hieraus und aus der nur dem Elementarunterricht dienenden Zielsetzung der Philosophischen Fakultäten im 18. Jahrhundert erklärt sich die große Abneigung von Gauss, für die wenig interessierten und ungenügend vorgebildeten Studenten Vorlesungen zu halten. Wirklich Begabten hat er in individueller Förderung stets geholfen. Vielleicht erklärt dies, daß seine persönlichen Schüler, von denen als Beispiele nur Encke, Möbius, v. Staudt, Dedekind, Listing und Minding genannt seien und denen man auf Grund seines Habilitationsvortrages wohl auch den von Dirichlet herkommenden Riemann zurechnen darf, zu einem nicht unerheblichen Teil nicht Arbeiten ihres Meisters fortsetzten, sondern eigene Wege gingen.

Kennzeichnend für die wissenschaftliche Arbeit von Gauss ist das Streben nach Ausschaltung jeder vermeidbaren Unsicherheit in der Wissenschaft und jeder wenn auch unbewußten subjektiven Verfälschung der Ergebnisse. Dieses Ziel eines Höchstmaßes von Objektivität erforderte die Abwendung von den laxen Beweismethoden des 18. Jahrunderts in der Mathematik und die Rückkehr zum « rigor antiquus », zu den strengen Beweisverfahren der antiken Mathematiker. In der Physik und den anderen messenden Wissenschaften führte es zwangsläufig zur Ablehnung willkürlicher Auswahlen von Messungen in einer Messreihe und zum Grundsatz der Gleichberechtigung aller ihr angehörenden Messungen. Dazu kam als weitere Forderung, daß die Grundsätze zur Bearbeitung einer Meßreihe schon vor der Reduktion feststehen sollten. Ebenso gehört hierher die sorgfältige Untersuchung der Instrumentalfehler und die Entwicklung von Meßverfahren, bei denen sie möglichst unschädlich gemacht werden. Die leidenschaftliche Wahrheitsliebe von Gauss hat dieses Programm nicht nur aufgestellt, sondern wesentlich zu seiner Verwirklichung beigetragen. An seine Arbeiten stellte Gauss nicht nur inhaltlich, sondern auch in der Form höchste Ansprüche. Das führte leider dazu, daß in der veröffentlichten Gestalt meist nicht mehr zu erkennen ist, wie die Ergebnisse gefunden wurden. Daraus ergeben sich oft Schwierigkeiten für ein wirkliches Verständnis.

Zu den ersten, schon 1792 beginnenden mathematischen Betätigungen von Gauss gehört seine Beschäftigung mit dem Parallelenaxiom. Wenn er schon als fünfzehnjähriger an eine Geometrie denkt, « die stattfinden müßte, wenn die Euklidische Geometrie nicht die wahre ist », so hat dieser Gedanke offenbar zu keinen unmittelbaren Konsequenzen geführt. Für die weitere Entwicklung ist die Überlieferung sehr lückenhaft, nur die Briefe an seinen Studienfreund Wolfgang Bolyai, der sich ebenfalls mit dem Parallelenaxiom beschäftigte, lassen einige Schlüsse zu. Gauss scheint danach wie verschiedene Vorgänger die Geometrie, die sich bei Verwerfung des Parallelenaxioms ergibt,

entwickelt zu haben, in der Hoffnung, auf einen Widerspruch zu stoßen. Dabei setzte sich bei ihm im Laufe einer von Rückschlägen nicht freien Entwicklung und offenbar gegen starkes inneres Widerstreben die Überzeugung von der Unmöglichkeit eines Beweises durch. Wie schwierig dieser Bruch mit der Überlieferung war, erhellt vielleicht am besten daraus, daß ein so hervorragender Mathematiker wie LEGENDRE nie dahin gelangte. Etwa 1816 war ein vorläufiger Abschluß erreicht, die « nichteuklidische Geometrie » war entstanden. Daß von den beiden denkbaren nichteuklidischen Geometrien, die wir heute als elliptische und hyperbolische Geometrie bezeichnen, damals und ebenso bei den Wiederentdeckern zunächst nur die hyperbolische Geometrie entstand, hängt damit zusammen, daß die elliptische Geometrie durch die Beibehaltung des eudoxisch-archimedischen Axioms (« eine Größe kann so oft vervielfältigt werden, daß sie jede andere gleichartige übertrifft ») ausgeschlossen wurde. GAUSS hielt mit seinen Erkenntnissen gerade hier ängstlich zurück und legte seinen Korrespondenten Stillschweigen auf, weil Andeutungen in einer Rezension ihm ihn verstimmende Angriffe eingetragen hatten. Es waren allerdings Teile der neuen Geometrie so denkungewohnt, daß dies nicht verwundern darf. Das Nichtvorhandensein ähnlicher Figuren und die Existenz einer ausgezeichneten Länge gehören dahin. In der Folgezeit hat sich GAUSS während seiner Beschäftigung mit der Flächentheorie wieder mit der nichteuklidischen Geometrie befaßt, wie der Briefwechsel zeigt. Er brach seine Untersuchungen aber fast völlig ab, als die nichteuklidische Geometrie unabhängig zweimal neu entdeckt wurde. Beginnend 1823 hatte JOAHNN BOLYAI, der Sohn von GAUSS Jugendfreund, damals Pionierleutnant in Temesvar, die nichteuklidische Geometrie neu entwickelt und sie 1831 veröffentlicht. GAUSS berichtet darüber seinem Schüler GERLING und schreibt: « Ich halte diesen jungen Geometer v. Bolyai für ein Genie erster Größe. » Ebenfalls 1823 begannen die Untersuchungen des russischen Mathematikers NIKOLAI IWANOWITSCH LOBATSCHEFSKIJ, Professor in Kasan. Seine ersten russischen Publikationen blieben in Westeuropa, wenn sie überhaupt dahin gelangten, völlig unbekannt. Erst die zweite deutsch geschriebene Arbeit von 1840 wurde infolge einer abfälligen Besprechung von GAUSS beachtet. GAUSS, der noch über sechzigjährig die russische Sprache erlernt hatte, verschaffte sich nun auch die russischen Abhandlungen LOBATSCHEFSKIJS. Die Ernennung des russischen Gelehrten zum Korrespondierenden Mitglied der Gesellschaft der Wissenschaften in Göttingen ist ein Zeichen der Wertschätzung, die GAUSS ihm entgegenbrachte.

Allgemeiner bekannt wurden die Leistungen von BOLYAI und LOBATSCHEFSKIJ erst durch die unermödliche Arbeit und die Übersetzungen von HOÜEL, der für beide noch die gewaltige Autorität von GAUSS durch Übersetzung von Teilen an SCHUMACHER gerichteter Briefe in die Wagschale warf. Es fehlt hier der Raum, die Bedeutung der Entdeckung der nichteuklidischen

Geometrie für die geometrische Forschung auch nur anzudeuten. Daher beschränken wir uns auf einen Hinweis: Die Geometrie EUKLIDS war der Idee nach die Lehre von der Ordnung unseres Erfahrungsraumes, also im Grunde ein Stück Physik. Nun stellte sich heraus, daß die neue Geometrie bei geeigneter Verfügung über die in ihr auftretende absolute Konstante dasselbe leistete, da die Abweichungen unter die Schranke der Meßgenauigkeit fielen. Der philosophische Anspruch auf den a priori-Charakter der euklidischen Geometrie wurde damit unhaltbar. Die Weiterentwicklung führte schließlich dazu, daß die Lehre vom Erfahrungsraum, soweit sie nicht zur Domäne der Psychologie gehört, unzweifelhaft ein Teil der Physik ist und bleiben wird.

Ebenfalls seit 1792 beschäftigte sich GAUSS mit drei Themen, die, scheinbar wesensfremd, im Laufe der Zeit zu einer großartigen Einheit zusammenwuchsen. Das erste ist das sog. « arithmetisch-geometrische Mittel »: gegeben seien zwei beliebige positive Zahlen a und b. Man bildet die beiden Zahlenfolgen

$$a_1 = \frac{a+b}{2}, \qquad a_2 = \frac{a_1+b_1}{2}, \quad \dots, \qquad a_{n+1} = \frac{a_n+b_n}{2}, \quad \dots,$$

$$b_1 = \sqrt{ab}, \qquad b_2 = \sqrt{a_1 b_1}, \quad \dots, \qquad b_{n+1} = \sqrt{a_n b_n}, \quad \dots.$$

Es ist leicht zu zeigen, daß diese beiden Folgen ein-und demselben Limes $\mathcal{M}(a, b)$, dem sog. arithmetisch-geometrischen Mittel zustreben. Nach dieser Definition ist übrigens $\mathcal{M}(a, b) = \mathcal{M}(a_1, b_1) = \dots$. Das zweite Thema sind Potenzreihen der Gestalt $\sum c_n x^{n^2}$ und verwandte Reihen. Zu diesen gehören die Reihen, die man heute Thetanullwerte nennt. Das dritte Thema ist die Untersuchung des Integrals für die Bogenlänge der Lemniskate $\int_0^x \mathrm{d}x/(\sqrt{1-x^4})$. GAUSS faßt dieses Integral als Verallgemeinerung des Integrals für den Arcussinus auf. Von dieser Analogie geleitet, betrachtet er die obere Grenze des Integrals als Funktion des Integralwertes. Diese von ihm Sinus lemniscaticus genannte Funktion ist die erste elliptische Funktion nach heutiger Bezeichnung. Für die Weiterentwicklung entscheidend war, daß GAUSS schon damals von den komplexen Zahlen konsequent Gebrauch machte. Die bekannte geometrische Veranschaulichung dieser Zahlen fand er erst 1799 und veröffentlichte sie erst 1831.

Seine Beschäftigung mit den sog. Kreisteilungsgleichungen führte GAUSS dazu, auch die Gleichung für die Teilung des Lemniskatenumfangs in n gleiche Teile aufzustellen. Zu seiner Überraschung war diese Gleichung nicht, wie erwartet, vom Grad n, sondern vom Grad n^2. Schon 1797 findet er, « *warum* » dies so ist. Diese lakonische Feststellung kann nichts anderes bedeuten, als daß er damals die zweite, imaginäre, Periode des Sinus lemniscaticus entdeckt

haben muß. Sein kurzer Hinweis in Artikel 335 seiner « Disquisitiones arithmeticae », daß sich seine Untersuchungen zur Kreisteilung u.a. ausdehnen lassen von den Kreisfunktionen auf die von dem Integral $\int_0^x \mathrm{d}x/\sqrt{1-x^4}$ abhängenden Funktionen, war übrigens das einzige, was GAUSS über diese Untersuchungen berichtete. Dieser Hinweis hat aber Geschichte gemacht, führte er doch ABEL und JACOBI zur Neuentdeckung der Theorie der elliptischen Funktionen.

Inzwischen war auch die Theorie des arithmetisch-geometrischen Mittels weiter gefördert. GAUSS hatte die Potenzreihenentwicklung von $\mathcal{M}(1+x, 1-x) = \mathcal{M}(1, \sqrt{1-x^2})$ versucht, war aber dann zur Entwicklung der Reziproken übergegangen, bei der sich ein überraschend einfaches Koeffizientengesetz ergab. 1799 bemerkte er zufällig, daß die Dezimalbrüche für $1/\mathcal{M}(1, \sqrt{2})$ und für

$$\frac{\tilde{\omega}}{\pi} = \frac{2}{\pi}\int_0^1 \frac{\mathrm{d}x}{\sqrt{1-x^4}}$$

(dieses $\tilde{\omega}$ ist die reelle Halbperiode des Sinus lemniscaticus) auf 11 Stellen übereinstimmten. Das führte ihn zur Entdeckung, daß

$$\frac{1}{\mathcal{M}(1, \sqrt{1-\nu^2})} = \frac{2}{\pi}\int_0^1 \frac{\mathrm{d}x}{\sqrt{(1-x^2)(1-\nu^2x^2)}}$$

sein mußte. Offenbar durch Unklarheiten über die Integration mehrdeutiger Funktionen im Komplexen veranlaßt, wendet sich GAUSS nicht der Umkehrung des allgemeinen elliptischen Integrals erster Gattung zu, sondern verallgemeinert formal ältere Ergebnisse über den $\sin \mathrm{lemn}\, \varphi$. Diesen hatte er in eine nach den Sinus der ungeraden Vielfachen von φ fortschreitende Reihe entwickelt und ihn außerdem als Quotienten zweier Thetareihen dargestellt. Dazu kamen für $\sqrt{\tilde{\omega}/\pi}$ zwei Darstellungen durch Thetanullwerte. GAUSS gibt nun ein μ vor und setzt

$$\frac{\pi}{\mathcal{M}(1, \sqrt{1+\mu^2})} = \tilde{\omega}\,, \qquad \frac{\pi}{\mu\mathcal{M}(1, \sqrt{1+1/\mu^2})} = \tilde{\omega}'\,.$$

Der « Sinus lemniscaticus universalissime acceptus » $\mathrm{Sl}\varphi$ ist dann der Quotient zweier mit den Argumenten $\varphi(\pi/\tilde{\omega})$ und $x = \exp[-\pi(\tilde{\omega}'/\tilde{\omega})]$ (dieses x ist nichts anderes als das Jacobische q) gebildeter Thetareihen. Es ist unmittelbar ersichtlich, daß dieses $\mathrm{Sl}\varphi$ die beiden Perioden $2\tilde{\omega}$ und $2\tilde{\omega}'i$ hat. Setzt man $\sin u = \mathrm{Sl}\varphi$, so folgt $\mathrm{d}u/\mathrm{d}\varphi = \sqrt{1+\mu^2 \sin^2 u}$ und damit, daß $\mathrm{Sl}\varphi$ die Umkehr-

funktion von

$$q \int_0^S \frac{\mathrm{d}S}{\sqrt{(1-S^2)(1+\mu^2 S^2)}}$$

ist. Dieses Ergebnis ist nur scheinbar befriedigend. Denn geht man von der Funktion $\mathrm{Sl}\,\varphi$ aus, so kann man $\tilde{\omega}$ und $\tilde{\omega}'$ oder auch x beliebig vorgeben mit der einen, für die Reihenkonvergenz nötigen Einschränkung, daß der Realteil, von $\tilde{\omega}'/\tilde{\omega}$ positiv, d.h. $|x|<1$ ist. $\tilde{\omega}$ erhält man aus einer Darstellung von $\sqrt{\tilde{\omega}/\pi}$ als Thetanullwert, und auch μ läßt sich durch Thetanullwerte berechnen. Vom Integral kann man nicht gut ausgehen, da μ irgend eine komplexe Zahl sein kann, die arithmetisch-geometrischen Mittel, durch welche $\tilde{\omega}$ und $\tilde{\omega}'$ erklärt sind, sich also nicht bilden lassen, weil nicht feststeht, wie die hierzu nötigen Wurzelziehungen gemeint sind. Aber selbst wenn das geklärt ist, bleibt fraglich, ob der Realteil von $\tilde{\omega}'/\tilde{\omega}$, wie nötig, positiv ausfällt.

Gauss läßt zunächst bei der Berechnung der $\mathcal{M}(a, b)$ definierenden Folgen a_j, b_j bei jedem Schritt die Wahl zwischen den beiden möglichen Quadratwurzeln frei. Dann erhält man zu den Ausgangsdaten a, b unendlich viele verschiedene $\mathcal{M}(a, b) = a\mathcal{M}(1, b/a)$, die nicht Null sind (die verschwindenden sind für uns ohne Interesse). Damit wird $\mathcal{M}(1, k')$ eine unendlich vieldeutige Funktion von k'. Die Reihen für die Thetanullwerte erlauben nun, $\mathcal{M}(1, k')$ zu uniformisieren, d.h. sowohl k' als auch $\mathcal{M}(1, k')$, als eindeutige Funktionen eines Parameters t anzugeben. Liefert ein bestimmtes t einen Wert von k', so hat man zunächst sämtliche andere t, welche denselben Wert von k' liefern, zu ermitteln. Die Beantwortung dieser Frage gelingt Gauss, da ihm gewisse Transformationseigenschaften der Thetanullwerte bekannt sind. Diese neuen t gehen aus dem ursprünglichen t durch gewisse gebrochene lineare Substitutionen hervor, die zusammen eine Gruppe bilden. Diese ist im Grunde nichts anderes als eine sog. Hauptkongruenzuntergruppe der bekannten Modulgruppe. $k' = k'(t)$ ist damit als Modulfunktion erkannt.

Gauss hat übrigens, wie aus Notizen hervorgeht, auch die heute « absolute Invariante » genannte, zur vollen Modulgruppe gehörige Modulfunktion in Betracht gezogen. Mehr ist darüber nicht bekannt, doch haben gerade diese Notizen 1877 Dedekind zur Einführung und Untersuchung dieser, von ihm « Valenz » genannten Funktion veranlaßt. Die Uniformisierung des arithmetisch-geometrischen Mittels mittels des Parameters t wird nun dazu benutzt, um solche $\tilde{\omega}$ und $\tilde{\omega}'$ als « zusammengehörig » zu erklären, welche zum gleichen t gehören. Es folgt dann, daß $t = \tilde{\omega}'/\tilde{\omega}$ ist, allerdings unter der Voraussetzung, daß es ein t gibt, welches das gegebene μ liefert. Gauss hat 1805 die Modulsubstitutionen geometrisch gedeutet und den Diskontinuitätsbereich der Gruppe in der t-Ebene wirklich aufgezeichnet. In diesem liegen nur t mit positivem reellen Teil. Man kann durch Abbildung des Randes des Diskon-

tinuitätsbereiches dessen Abbildung in die k'-Ebene erschließen und daraus entnehmen, daß k'^2 im Diskontinuitätsbereich der Gruppe jeden vorgeschriebenen Wert annimmt, und zwar jeden genau einmal. Damit sind dann alle Fragen beantwortet.

Die gegebene Skizze der Gaußschen Arbeiten zur Theorie der elliptischen Funktionen ist sehr unvollständig und vereinfacht den komplizierten Entwicklungsgang ganz erheblich. Daß wir den Nachlaß von GAUSS hier verstehen, verdanken wir vor allem der hingebenden Arbeit von L. SCHLESINGER. Sein Schüler GEPPERT hat dann den Nachlaß von GAUSS zur Theorie der elliptischen Funktionen mit verhältnismäßig geringen Ergänzungen zu einem lesbaren Ganzen gestaltet [5].

In der gewöhnlich unter dem abgekürzten Titel « Determinatio attractionis annuli elliptici » (Werke III) zitierten Abhandlung über die säkularen Störungen, die ein Planet auf einen anderen ausübt, hat GAUSS einige seiner Ergebnisse über elliptische Integrale (nicht über elliptische Funktionen) und ihre Beziehungen zum arithmetisch-geometrischen Mittel mitgeteilt. Zur Bestimmung dieser Störungen genügt es, den Planeten zu ersetzen durch seine Bahnellipse, über die seine Masse so verteilt ist, daß die auf einem Bogen liegende Masse zur Gesamtmasse sich verhält, wie die zur Durchlaufung des Bogens nötige Zeit zur Gesamtumlaufszeit. Die Anziehung, die ein solcher Ring nach dem Newtonschen Gesetz ausübt, kann durch elliptische Integrale ausgedrückt werden.

Seine Arbeiten über elliptische Funktionen veranlaßten GAUSS auch, sich mit linearen Differentialgleichungen zweiter Ordnung zu beschäftigen. Von seinen Untersuchungen hat er nur einen ersten Teil unter dem Titel « Disquisitiones generales circa seriem infinitam

$$1 + \frac{\alpha\beta}{1\cdot\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot2\cdot\gamma(\gamma+1)}x^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot2\cdot3\cdot\gamma(\gamma+1)(\gamma+2)}x^3 + \ldots .»$$

(Werke III) 1813 veröffentlicht. Die Summe $F(\alpha, \beta, \gamma, x)$ dieser « hypergeometrischen Reihe » ist schon deshalb wichtig, weil eine Unzahl elementarer und höherer Funktionen, so z.B. die Perioden elliptischer Integrale als Funktionen des Moduls, durch Spezialisierung der Parameter α, β, γ aus ihr erhalten werden können. Daß GAUSS hier vom Komplexen keinen Gebrauch macht, ist merkwürdig, da er in der geplanten Fortsetzung es doch hätte tun müssen. In seiner Abhandlung verwendet GAUSS wohl als erster konsequent den modernen Begriff der Reihenkonvergenz. Für EULER war der Ausgangspunkt stets eine Funktion, die in eine Reihe entwickelt wurde. Operiert EULER mit solchen Reihen auch für Argumente, für welche sie divergieren, so läßt sich das

eigentlich immer, wenn auch oft recht mühsam, durch ein Operieren mit den Funktionen ersetzen. Bei GAUSS ist aber der Ausgangspunkt die Reihe, welche durch ihre Summe erst die Funktion definiert. Deshalb die Forderung der Konvergenz. Was GAUSS fehlt, ist das Prinzip der analytischen Fortsetzung. Daher wählt er als Ausweg die Erklärung seiner Funktionen außerhalb des Konvergenzkreises als Lösungen der linearen Differentialgleichung, der sie genügen. Auf den reichen Inhalt der Arbeit und der geplanten Fortsetzung näher einzugehen, verbietet der Raum, doch sei wenigstens darauf hingewiesen, daß sie in Fortsetzung von Untersuchungen EULERS eine Theorie der Gammafunktion (GAUSS nennt sie Π-Funktion, $\Pi(x) = \Gamma(x+1)$) enthält.

Aus dem Briefwechsel geht hervor, daß GAUSS bekannt war, daß das Integral einer eindeutigen Funktion im Komplexen eine mehrdeutige Funktion der oberen Grenze sein kann. Insbesondere weiß er, daß $\int \mathrm{d}x/x$ erstreckt über einen geschlossenen, $z = 0$ n-mal umschlingenden Weg gleich $2n\pi i$ ist. Ferner ist ihm bekannt, daß, modern ausgedrückt, das Integral einer eindeutigen analytischen Funktion in einem einfachzusammenhängenden Regularitätsbereich vom Integrationsweg unabhängig ist. GAUSS hat damit wesentliche Ergebnisse von CAUCHY um Jahrzehnte vor diesem besessen.

GAUSS' Dissertation vom Jahre 1799 « Demonstratio nova theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse » (Werke III) ist dem sog. Fundamentalsatz der Algebra gewidmet. Es sei

$$P(z) = a_0 + a_1 z + a_2 z^2 + \ldots + a_{n-1} z^{n-1} + z^n ,$$

ein Polynom mit reellen Koeffizienten. Wir beweisen heute, daß es n reelle oder komplexe (nicht notwendig verschiedene) Zahlen $z_1, z_2, \ldots, z_n$ gibt, so daß

$$P(z) = (z - z_1)(z - z_2) \ldots (z - z_n)$$

ist. Da zu jedem komplexen z_j auch das konjugiert komplexe auftritt, kann man die nichtreellen Faktoren paarweise zu reellen Faktoren zweiten Grades zusammenfassen. Daher die Formulierung des Satzes im Titel der Dissertation. Der Beweis von GAUSS verläuft, so weit dies möglich ist, ganz im Reellen. Vorausgeschickt ist eine eingehende Kritik der älteren Beweise. Wie das « Demonstratio nova » anzeigt, hält GAUSS diese älteren Beweise nicht für falsch, sondern für ergänzungsbedürftig. Setzt man $z = x+iy = r(\cos\varphi + i\sin\varphi)$, so wird $P(z) = Q(x, y) + iR(x, y)$. Die Nullstellen von $P(z)$ findet man also als Schnitte der Kurve $Q(x, y) = 0$ mit der Kurve $R(x, y) = 0$. Da man leicht einsieht, daß für sehr große r die Kurve $Q = 0$ näherungsweise zusammenfällt mit den

Strahlen durch $x = y = 0$, für die $\cos n\varphi = 0$ ist, die Kurve $R = 0$ aber mit den Strahlen, für die $\sin n\varphi = 0$ ist, und da keine der beiden Kurven im Endlichen frei enden kann, so schließt man geometrisch leicht auf das Vorhandensein von Schnittpunkten. GAUSS hat für denselben Satz später noch drei weitere Beweise gegeben (Werke III). Der zweite Beweis ist eine Präzisierung des in der Dissertation besprochenen Beweisversuches von EULER und FONCENEX, der vierte gibt eine Neuformung des Beweises der Dissertation. Besonders interessant ist der dritte Beweis. Er ist im Grunde eine Anwendung des Residuensatzes auf die logarithmische Ableitung von $P(z)$. Verschleiert wird das dadurch, daß das Kurvenintegral mit Hilfe des Gaußschen Integralsatzes aus der Potentialtheorie in ein Flächenintegral umgeformt ist.

Den Ruhm von GAUSS haben seine 1791 beginnenden Untersuchungen zur Zahlentheorie begründet, die in der Hauptsache in den 1801 erschienenen « Disquisitiones arithmeticae » (Werke I) enthalten sind. Dieses Werk umfaßt drei deutlich gesonderte Teile. Der erste, der Abschnitt 1 bis 4, beruht auf der Darstellung ganzer Zahlen als Produkte von Primzahlen, gehört also zur « multiplikativen » Zahlentheorie, der zweite, die Abschnitte 5 und 6, behandelt die quadratischen Formen und gehört in die additive Zahlentheorie. Die spätere Entwicklung hat daraus die Zahlentheorie der quadratischen Zahlkörper entwickelt, d.h. die Einordnung in die multiplikative Zahlentheorie vollzogen. Der dritte Teil, Abschnitt 7, gibt die Anwendung der Zahlentheorie auf die Kreisteilungsgleichung. Wie schon oben gesagt wurde, war GAUSS bekannt, daß sich diese Untersuchungen weiter ausdehnen ließen. z.B. auf die Gleichungen für die Teilung der Lemniskate. Der erste Teil war im wesentlichen fertig, als GAUSS 1795 nach Göttingen kam. Er mußte dort entdecken, daß vieles, was er selbständig gefunden hatte, bereits FERMAT, EULER, LAGRANGE und LEGENDRE bekannt war, doch blieb genug übrig, was ihm allein gehörte. Der zweite Teil geht wesentlich auf Anregungen durch die Arbeiten von LAGRANGE zurück.

Ein wesentliches Hilfsmittel ist für GAUSS der von ihm eingeführte Begriff der Kongruenz zweier ganzer Zahlen a und b nach einem ganzzahligen und von Null verschiedenen « Modul » m: $a \equiv b \bmod m$ (gelesen « a kongruent b modulo m ») bedeutet, daß $a - b$ durch m ohne Rest teilbar ist. Die Kongruenz hat, und das ist wesentlich, die charakteristischen Eigenschaften einer Gleichheitsbeziehung. So kommt man zu einer Analogie zwischen Kongruenzen und Gleichungen. Z.B. tritt neben die Gleichung $x^2 = D$ die Aufgabe, bei gegebenem und durch m nicht teilbaren ganzen D die ganzzahligen x zu finden, für welche $x^2 \equiv D \bmod m$ ist. Aus dem Inhalt des ersten Teiles der Disquisitiones arithmeticae greifen wir zwei besonders wichtige Ergebnisse heraus. Ist p eine Primzahl und a kein Vielfaches von p, so gibt es einen kleinsten positiven Exponenten t, für den $a^t \equiv 1 \bmod p$ ist. Man sagt dann: « a gehört

zum Exponenten t ». Dieses t ist entweder gleich $p-1$ oder ein Teiler von $p-1$. Eine Folge davon ist zunächst der « kleine Fermatsche Satz »: « Stets ist $a^{p-1} \equiv 1 \bmod p$ ». Zu jedem Teiler t von $p-1$, insbesondere also auch zum Exponenten $p-1$ gehören dann wirklich Zahlen a. Die zu p 1 gehörigen nennt man « primitive Wurzeln mod p ». Ist g eine solche, so ist jede durch p nicht teilbare Zahl einer der Potenzen $1 = g^0, g^1, g^2, \ldots, g^{p-2} \bmod p$ kongruent. Die Wichtigkeit dieses Ergebnisses zeigt seine Verwendung im dritten Teil der Disquisitiones.

Die Frage nach der Lösbarkeit der Kongruenz $x^2 \equiv D \bmod m$ kann man auf den Fall zurückführen, daß m eine ungerade Primzahl p und D nicht durch p teilbar ist. Ist in diesem Fall die Kongruenz lösbar, so nennt man D einen quadratischen Rest mod p und setzt mit LEGENDRE (GAUSS benutzt diese Symbolik nicht) $(D/p) = +1$, ist sie unlösbar, so ist D ein quadratischer Nichtrest mod p, und man setzt $(D/p) = -1$. Man kann nun eine doppelte Frage stellen: Welche D sind bei gegebenem p quadratische Reste und welche Nichtreste? Für welche p ist eine gegebenes D quadratischer Rest, für welche Nichtrest? Für $D = -1$ und $D = 2$ war die zweite Frage schon von EULER beantwortet. EULER, LEGENDRE und GAUSS, dieser durch numerische Induktion, hatten unabhängig von einander den Satz entdeckt, der in LEGENDRES Formulierung lautet; « Sind p und q verschiedene ungerade Primzahlen, so ist stets

$$\left(\frac{p}{q}\right) \cdot \left(\frac{q}{p}\right) = (-1)^{(p-1)/2 \cdot (q-1)/2}.$$

GAUSS nennt diesen Satz den Fundamentalsatz (theorema fundamentale) der Theorie der quadratischen Reste, heute nennt man ihn nach der gegebenen Formulierung mittels des Legendreschen Symbols meist das quadratische Reziprozitätsgesetz. Er besagt, ausführlich ausgesprochen: ist mindestens eine der Primzahlen p und q von der Form $4n+1$, so sind die Kongruenzen

$$x^2 \equiv q \bmod p \quad \text{und} \quad y^2 \equiv p \bmod q$$

beide lösbar oder beide unlösbar, sind dagegen beide Primzahlen von der Gestalt $4n+3$, so ist eine der Kongruenzen lösbar, die andere unlösbar. LEGENDRE hatte den Beweis dieses Satzes versucht, mußte dabei aber eine ihm selbst bekannte Lücke lassen, die erst DIRICHLET auszufüllen vermochte. GAUSS lieferte, ohne von LEGENDRE zu wissen, 1796 den ersten noch recht mühsamen Beweis, der in den Disquisitiones veröffentlicht wurde. Im zweiten Teil der Disquisitiones gab er einen zweiten Beweis, dem noch sechs weitere folgten. Das Reziprozitätsgesetz hat seine überragende Bedeutung immer wieder

gezeigt, sei es als Beweishilfsmittel, sei es dadurch, daß jeder neue Beweis für diesen Satz zu neuen Fragestellungen führte, sei es durch die von ihm ausgehenden Versuche, analoge Aussagen etwa für kubische oder biquadratische Reste zu finden.

Der zweite Teil der Disquisitiones befaßt sich mit der Darstellung ganzer Zahlen n durch sog. (binäre) quadratische Formen, d.h. mit Darstellungen der Gestalt $n = ax^2 + 2bxy + cy^2$. Dabei sollen a, b, c fest vorgegebene ganze Zahlen und auch x und y ganzzahlig sein. Sind dabei x und y teilerfremd, so nennt man die Darstellung von n eine eigentliche Darstellung. Auf solche eigentliche Darstellungen darf man sich beschränken. Geht die Form $ax^2 + 2bxy + cy^2$ durch eine lineare Substitution $x = \alpha x' + \beta y'$, $y = \gamma x' + \delta y'$ mit ganzzahligen Koeffizienten α, β, γ, δ in die Form $a'x'^2 + 2b'x'y' + c'y'^2$ über und läßt sich diese durch eine ebensolche Substitution wieder in die ursprüngliche zurückverwandeln, so nennt man beide Formen äquivalent. Äquivalente Formen stellen offenbar dieselben Zahlen n dar. Für die linearen Substitutionen, die äquivalente Formen in einander überführen, muß notwendig $\alpha\delta - \beta\gamma = \pm 1$ sein. Steht hier rechts $+1$, so nennt man die Formen eigentlich äquivalent (auf diesen Fall darf man sich beschränken), sonst uneigentlich äquivalent. Die Substitutionen mit $\alpha\delta - \beta\gamma = +1$ sind aber nichts anderes, als die bei den elliptischen Funktionen auftretenden Modulsubstitutionen. Es zeigt sich nun, daß bei äquivalenten Formen $b^2 - ac = b'^2 - a'c'$ ist. Die Gleichheit dieser Zahlen, die Gauss die Determinante D der Form nennt, heute nennt man sie meist Diskriminante, ist somit notwendig für die Äquivalenz. Es ist leicht an Beispielen zu sehen, daß nicht alle quadratischen Formen derselben Determinante D äquivalent zu zein brauchen und daß nicht jede Zahl durch eine solche Form darstellbar sein wird. Für die Darstellbarkeit einer Zahl n durch eine Form mit der Determinante D erweist sich als notwendig die Lösbarkeit der Kongruenz $z^2 \equiv D \bmod n$. Damit ist die Verbindung zum Reziprozitätsgesetz hergestellt. Gauss zeichnet unter allen Formen mit der Determinante D gewisse, deren Koeffizienten durch Ungleichheiten beschränkt sind, aus. Diese Formen nennt er reduzierte. Zu jedem D gehören nur endlich viele verschiedene reduzierte Formen und jede Form mit der Determinante D ist einer dieser reduzierten Formen äquivalent (wobei es möglich ist, daß auch reduzierte Formen unter sich äquivalent sind). Faßt man alle zu einer bestimmten Form der Determinante D äquivalenten Formen in eine Klasse zusammen, so gibt es demnach nur endlich viele Klassen für ein gegebenes D. Da die zur Klasseneinteilung der quadratischen Formen führenden Überlegungen nicht rein formal sind, sondern sich wesentlich auf innere Eigenschaften der Formen gründen, darf man erwarten, daß sie auch zu zahlentheoretischen Sätzen führen. In der Tat folgen jetzt von Fermat angegebene und von Euler und Lagrange bewiesene Sätze über die Dar-

stellung gewisser Primzahlen durch spezielle quadratische Formen, denen GAUSS weitere der gleichen Art anfügt. Vor allem aber folgen die weiteren Entwicklungen des zweiten Abschnitts der Disquisitiones, deren Darstellung sich wohl einer elementaren Darlegung entzieht.

In den « Disquisitiones arithmeticae » kommen nirgends geometrische Überlegungen vor, nirgends werden geometrische Beziehungen auch nur angedeutet. Nun legt schon die Bezeichnung « kongruent » die Vermutung nahe, daß solche Beziehungen bestehen müssen, zur Gewißheit wird sie durch das Auftreten der Modulsubstitutionen bei der Überführung einer quadratischen Form in äquivalente. In der Theorie der elliptischen Funktionen dienen diese Substitutionen dazu, um von einem Periodengitter zu einem äquivalenten, d.h. gleichwertigen, überzugehen. Für die Darlegung des Zusammenhangs beschränken wir uns auf den Spezialfall einer positiv definiten quadratischen Form $ax^2+2bxy+cy^2$, d.h. einer Form, die außer für $x = y = 0$ nur positive Werte annimmt. Es ist dann $a > 0$, $b > 0$ und $D < 0$. Jeder solchen Form ordnet man umkehrbar eindeutig ein Parallelogrammgitter folgendermassen zu: In einem XY-Koordinatensystem mit dem Achsenwinkel φ, der durch $\cos\varphi = b/\sqrt{ac}$ bestimmt ist, zieht man durch die Punkte $X = x\sqrt{a}$ (x ganz) der X-Achse die Parallelen zur Y-Achse und ebenso durch die Punkte $Y = y\sqrt{c}$ (y ganz) der Y-Achse die Parallelen zur X-Achse. Die Ecken $(x\sqrt{a},\ y\sqrt{c})$ dieses Parallelogrammgitters nennen wir Gitterpunkte. Betrachten wir nur diese, so gibt es unendlich viele Parallelogrammgitter, die zu denselben Gitterpunkten gehören, und zwar entsprechen alle diese den Formen, die zur Ausgangsform äquivalent sind. Stets haben die Maschen eines solchen Netzes die Fläche $\sqrt{-D}$. Die Parallelogrammgitter entsprechen so den einzelnen Formen, das System der Gitterpunkte aber der von diesen Formen gebildeten Klasse. Bei den (ternären) quadratischen Formen in drei Veränderlichen, die GAUSS ebenfalls heranzieht, kommt man ebenso auf Punktgitter im Raum. Diesen Zusammenhang hat GAUSS nur gelegentlich bei der Besprechung einer Arbeit eines seiner Schüler angegeben, und er hat dabei auch den Zusammenhang mit der Kristallographie hervorgehoben (Werke II, S. 188). Die Kongruenzen geben, wie man nun sofort sieht, die eindimensionalen Punktgitter.

1796 erschien GAUSS erste Veröffentlichung, seine Mitteilung, daß das regelmäßige 17-Eck mit Zirkel und Lineal allein konstruierbar ist (Werke X_1). Dies war seit den griechischen Mathematikern der erste Fortschritt in der Theorie der regelmäßigen Vielecke. Aus dieser Entdeckung ist wohl der dritte Teil der Disquisitiones entsprungen.

Der Teilung des Kreises in n gleiche Teile entspricht algebraisch die Lösung

der Gleichung $x^n - 1 = 0$. Da man in dieser Gleichung den Linearfaktor $x - 1$ abspalten kann, hat man nur die Gleichung (die wir künftig Kreisteilungsgleichung nennen werden)

$$x^{n-1} + x^{n-2} + \ldots + x + 1 = 0$$

zu betrachten. Dabei beschränken wir uns auf den Fall, daß n eine Primzahl ist. Die Lösungen dieser Gleichung sind die ganzen Potenzen von $e^{2\pi i/n}$, aber damit ist die Berechnung der Lösungen auf transzendente Hilfsmittel (unendliche Reihen) zurückgeführt. Das verdeckt die wesentliche Eigenschaft der Lösungen, algebraische Zahlen zu sein, und verhindert damit jede tiefere Einsicht in ihren Zahlcharakter. Daher gibt GAUSS eine systematische algebraische Auflösungstheorie für die Kreisteilungsgleichung. Zunächst zeigt er, daß die Gleichung irreduzibel ist, d.h., daß die linke Seite nicht als Produkt zweier Polynome mit rationalen Koeffizienten geschrieben werden kann. Es sei weiter r eine Lösung der Kreisteilungsgleichung. Dann ist $r^n = 1$. Gibt man weiter $n - 1$ ganze Zahlen α, die in irgend einer Reihenfolge den Zahlen $1, 2, \ldots, n-1$ nach dem Modul n kongruent sind, so sind die r^α sämtliche Lösungen der Kreisteilungsgleichung. Für die Zahlen α nimmt nun GAUSS die Potenzen $g^0 = 1, g^1, g^2, \ldots, g^{n-2}$ einer primitiven Wurzel g mod n. Wichtig ist, daß alles weitere von der Wahl von g unabhängig ist, d.h. daß man dasselbe Ergebnis erhält, wenn man von verschiedenen primitiven Wurzeln mod n ausgeht. $n - 1$ ist als gerade Zahl sicher zusammengesetzt. Es sei $n - 1 = e \cdot f$, wobei keiner der Faktoren gleich 1 ist. Dann ordnet GAUSS die Potenzen von r in e Systeme zu je f Lösungen, und zwar so, daß in jedem System die Exponenten von g (die Potenzen von g sind ja jetzt unsere α) eine arithmetische Progression mit der Differenz e bilden. Es seien $\eta_0, \eta_1, \ldots, \eta_{e-1}$ die Summen der Potenzen von r in jedem dieser Teilsysteme. Dann genügen die η einer Gleichung vom Grad e mit ganzen rationalen Koeffizienten. Ist f keine Primzahl, sondern zusammengesetz, $f = e' \cdot f'$, so betrachten wir die Potenzen von r, die etwa in η_0 vorkommen. Diese ordnen wir ebenso in e' Teilsysteme zu je f' solcher Lösungen, bei denen die Exponenten von g arithmetische Progressionen mit der Differenz $e \cdot e'$ bilden. Bilden wir dann in jedem System die Summen der r-Potenzen, so genügen diese einer Gleichung vom Grad e', deren Koeffizienten linear von den η abhängen. Dieses Verfahren wiederholt man so lange wie möglich. Hat $n - 1$ die Primfaktorzerlegung $n - 1 = p_1^{\varrho_1} \cdot p_2^{\varrho_2} \ldots p_k^{\varrho_k}$, so wird auf diese Weise die Auflösung der Kreisteilungsgleichung zurückgeführt auf die Lösung von ϱ_1 Gleichungen vom Grade p_1, ϱ_2 Gleichungen vom Grade p_2, $\ldots$, ϱ_k Gleichungen vom Grade p_k. Die Konstruktion eines regulären Vielecks mit Zirkel und Lineal allein gelingt immer dann und nur dann, wenn ihm algebraisch eine Lösung der Kreisteilungsgleichung durch eine Kette quadratischer Gleichungen entspricht, d.h. wenn $n - 1$ eine Potenz von 2 ist. Es

ist leicht zu zeigen, daß, da n Primzahl sein soll, der Exponent von 2 wieder eine Potenz von zwei sein muß.

Von den späteren zahlentheoretischen Arbeiten von GAUSS erwähnen wir nur, daß er bei seinen Untersuchungen über biquadratische Reste (Werke II) aus dem Körper der rationalen Zahlen herausgeht und den heute sogenannten Gaußschen Zahlenkörper einführt. Dieser besteht aus allen komplexen Zahlen $a+bi$ mit rationalem a und b. Ganze Zahlen sind in diesem Körper die Zahlen mit ganzem rationalem a und b. Damit war die ursprüngliche Beschränkung der Zahlentheorie auf den rationalen Zahlkörper beseitigt und der Weg zur Fortentwicklung frei gemacht.

Im Nachlaß von GAUSS finden sich noch eine Reihe von Sätzen über sog. asymptotische Gesetze in der Zahlentheorie. Wir erwähnen nur ein Beispiel: Ist $\pi(x)$ die Anzahl der Primzahlen, die x nicht übertreffen, so gibt GAUSS an, daß

$$\pi(x) \sim \int_2^x \frac{\mathrm{d}t}{\log t}, \qquad \text{d.h.} \qquad \lim_{x\to\infty} \frac{1}{\pi(x)} \int_2^x \frac{\mathrm{d}t}{\log t} = 1$$

ist. Der Beweis dieses « Primzahlsatzes » gelang erst 1896 HADAMARD und DE LA VALLÉE POUSSIN. GAUSS selbst hat für diesen und andere hierhergehörige Sätze wohl keinen schlüssigen Beweis besessen.

* * *

In der Nacht des 1. Januar 1801 (nach astronomischer Datierung) beobachtete G. PIAZZI in Palermo ein Sternchen mit einer beträchtlichen Eigenbewegung. Dieses konnte er sechs Wochen lang messend verfolgen. Er war sich im Zweifel, ob es sich um einen Kometen oder einen neuen Planeten handelte. Die Mitteilung seiner Entdeckung langte erst nach Monaten in Paris an und fand dort wenig Beachtung. Noch viel später erreichte sie die deutschen Astronomen. Da es sich um ein wenig auffallendes Sternchen handelte, konnte man nur dann hoffen, es bei der nächsten Opposition wiederzufinden, wenn man seinen Ort wenigstens näherungsweise kannte. Um eine solche Angabe zu machen, mußte man die Bahn des neuen Planeten vorausberechnen. Nun trat die Aufgabe, für ein neues Mitglied unseres Planetensystems die Bahn zu ermitteln, nicht zum ersten Mal auf. 1781 hatte HERSCHEL den Uranus entdeckt. Da es sich dabei offenbar um einen großen Planeten handelte und man von diesen gewohnt war, daß die Bahnebene von der Ebene der Ekliptik wenig verschieden und die Exzentrizität der Bahn klein war, berechnete man für Uranus zuerst eine Kreisbahn in der Ekliptikebene. Diese

Näherung war so gut, daß man mit ihr eine große Anzahl älterer Beobachtungen, die sich über ein Jahrhundert verteilten, auffinden und zur Bahnverbesserung benutzen konnte. Für die zahlreichen nichtperiodischen Kometen war man gewohnt Parabelbahnen zu berechnen, d.h. die Exzentrizität gleich 1 zu setzen. Ein Verfahren zu einer Bestimmung einer Parabelbahn hatte schon I. Newton angegeben (« Philosophiae naturalis principia mathematica, Liber III Propositio XLI Problema XXI: Cometae in Parabola moti Trajectoriam ex datis tribus Observationibus determinare »). Olbers hatte, gestützt auf die Arbeiten von Lambert und anderen, ein neues besseres Verfahren angegeben, das im wesentlichen heute noch in Gebrauch ist. Halley hatte für die Kometen von 1682, 1607, 1531 und 1456 Parabelbahnen berechnet, die eine so auffallende Übereinstimmung der Elemente, d.h. der sechs die Bahn festlegenden Zahlen, aufwiesen, daß er schloß, es müsse sich um ein und denselben Kometen handeln. Damit war nicht nur der erste periodische Komet entdeckt, sondern für diesen auch die Umlaufszeit und damit die halbe Achse der Bahnellipse bekannt. D.h. er kannte ein sehr wichtiges Element der Bahn. Da er weiter schließen konnte, daß die von ihm berechneten Parabeln in der Nähe des Perihels die Bahnellipse gut annäherten, hatte er auch für die übrigen Elemente brauchbare Näherungen.

Im Falle des vom Piazzi entdeckten und Ceres genannten Planeten war nun für keines der sechs Elemente eine brauchbare Näherung bekannt. Zwar versuchte man hier auch die Berechnung einer Kreisbahn und einer Parabelbahn und sogar unter willkürlicher Annahme des Ortes des Perihels einer Ellipsenbahn. Aber alle diese Versuche hätten, wie sich hinterher herausstellte, zu keinem brauchbaren Ergebnis geführt. In diesem Stadium griff Gauss ein. Er stellte sich die Aufgabe, eine Bahn ohne jede willkürliche Annahme über die sechs Elemente auf Grund der Beobachtungen zu berechnen. Dazu müssen in der Regel drei Beobachtungen genügen, von denen jede Deklination und Rektaszension des Planeten zur Beobachtungszeit ergibt. Die Gleichungen, welche die Beobachtungen mit den Elementen verbinden, sind ziemlich verwickelt und können nur durch Näherungen gelöst werden. Dabei kommt es wesentlich auf eine geschickte Anordnung und Führung der Rechnung an. Außerdem liegen i.a. mehr als drei Beobachtungen vor. Selbst wenn man für die Auswahl von dreien unter allen eine vernünftige Regel hätte, so bliebe immer noch die Frage, wie man die überschüssigen Beobachtungen verwendet, um die aus den vorher ausgewählten Beobachtungen ermittelte Bahn zu verbessern. Gauss geht daher einen anderen Weg, bei dem die Gleichberechtigung aller Beobachtungen leitender Gedanke ist. Er benutzt dazu die von ihm entwickelte Ausgleichsrechnung, heute meist Methode der kleinsten Quadrate genannt. Über diese wird später noch einiges zu sagen sein. Mit ihrer Hilfe berechnet er aus allen Beobachtungen drei « Normalörter ». Diese legt er der Bahnbestimmung zu Grunde. Aus der Art, wie diese Normalörter gebildet

werden, folgt, daß bei ihnen die unvermeidlichen Beobachtungsfehler gut kompensiert sind. Dies war schon deshalb wichtig, weil bei der Ceres sich die Beobachtungen auf ein sehr kurzes Stück der Bahn verteilten. So gelang GAUSS die Lösung der selbstgestellten Aufgabe. Er berechnete auf Grund der von ihm ermittelten Elemente eine Ephemeride der Ceres, mit der dann die Wiederauffindung sofort gelang. Dieser auch dem Nichtfachmann verständliche Erfolg machte GAUSS weithin bekannt und berühmt.

Die nächsten Jahre verwendete GAUSS auf die Verbesserung seines Bahnbestimmungsverfahrens. Dieses hatte inzwischen erhöhtes Interesse dadurch gewonnen, daß in rascher Folge drei weitere kleine Planeten, Pallas, Juno und Vesta, entdeckt worden waren. Die zusammenfassende Veröffentlichung erschien erst 1809 als « Theoria motus corporum coelestium in sectionibus conicis solem ambientium » (Werke VII).

Die Ceres wurde nicht genau an dem vorausberechneten Ort wiedergefunden. Dies lag in erster Linie an der Nichtberücksichtigung der Störungen ihrer Bahn durch die großen Planeten. Für den zu erreichenden Zweck, die Wiederauffindung der Ceres, war es nicht nötig, diese Störungen zu ermitteln. Außerdem wäre das bei der Kürze der zur Verfügung stehenden Zeit und dem damaligen Stand der Theorie auch kaum möglich gewesen. Es war aber nur folgerichtig, daß GAUSS sich im Anschluß an seinen Erfolg mit den Störungen der Ceresbahn beschäftigte. Er beendete diese Arbeit aber nicht, sondern wandte sich, wohl veranlaßt durch ein Preisausschreiben der Pariser Académie des Sciences, der Ermittlung der Pallasstörungen zu. Die von ihm hier benutzten Methoden waren neuartig. Aus uns unbekannten Gründen brach er aber auch diese Rechnungen und Untersuchungen kurz vor ihrer Vollendung ab.

Bei der Übernahme der Sternwarte in Göttingen fand GAUSS kaum brauchbare Instrumente vor. Schon nach den ersten Neuanschaffungen begann eine rege Tätigkeit. Zunächst handelte es sich um die Bestimmung der Koordinaten der Sternwarte und dann um die Ermittlung wichtiger astronomischer Konstanten. Es mußten solche Merkwürdigkeiten aufgeklärt werden, wie die, daß die Messungen der Schiefe der Ekliptik nicht nur im Sommersolstitium wesentlich andere Werte ergaben, wie die Messungen im Wintersolstitium, sondern daß auch die Messungen verschiedener Astronomen hier keine Übereinstimmung zeigten.

Eine Ausnahme bildeten allein die Messungen von BESSEL und die von diesem reduzierten Messungen von BRADLEY. Dies führte zu einer eingehenden Untersuchung der möglichen Fehlerquellen in enger Zusammenarbeit mit BESSEL. Diese der Erhöhung der Genauigkeit astronomischer Messungen gewidmeten Untersuchungen setzten sich durch viele Jahre fort, jedoch ging die Führung bei der großen Belastung von GAUSS durch andere Arbeiten immer mehr an BESSEL über. Es ist aber fesselnd zu sehen, wie diese Untersuchungen

in engstem Zusammenhang stehen mit zahlreichen anderen, bei denen fast der ganze Umkreis der messenden Astronomie durchschritten wurde, von Untersuchungen über die Refraktion bis zu solchen über die Eigenbewegung des Sonnensystems. Von rein astronomischen Ergebnissen sei nur das schöne Verfahren zur gleichzeitigen Bestimmung von Polhöhe und Uhrkorrektion durch Feststellung der Uhrzeiten, in denen drei Sterne dieselbe Zenitdistanz erreichen, genannt. Der Bau der Gauss bei seiner Berufung in Aussicht gestellten neuen Sternwarte rückte nur langsam vorwärts, da die napoleonischen Kriege und die Mißwirtschaft des Hofes die Finanzkraft des Staates übermäßig erschöpften. Auch nach dem Wiener Kongress dauerte die Geldnot noch viele Jahre an. Schließlich aber entstand ein für die damalige Zeit gut ausgestattetes Institut. Lange Meßreihen, wie sie die Astronomen lieben und auch benötigen, lieferte zwar die neue Sternwarte nicht. Dafür haben aber zahlreiche bekannte Astronomen hier ihre Ausbildung erhalten.

1816 schlug Schumacher, der im Auftrag der dänischen Regierung an einer Gradmessung von der Nordspitze Jütlands bis an die Elbe arbeitete, Gauss die Fortsetzung dieser Messung nach Süden vor. Gauss ging darauf mit Freuden ein. Es mag dabei die Kritik, die Gauss an der französischen Gradmessung von 1792 schon deshalb übte, weil für die Abplattung der Erde nicht der aus den Messungen folgende Wert, sondern ein auf die Gradmessung in Peru zurückgehender genommen wurde, mitgespielt haben. Eng damit zusammen hängt die Erkenntnis, daß die Gestalt der Erde, die Gauss als erster als Niveaufläche der Schwerkraft erklärt, kein Ellipsoid sein könne. Gauss hatte, als er auf Schumachers Vorschlag einging, schon weiterreichende Pläne. Er wollte versuchen, seine Gradmessung mit den Vermessungsarbeiten in Bayern und Oesterreich in Verbindung zu bringen, um so einen Meridianbogen zu erhalten, der vom äußersten Norden Jütlands bis nach Oberitalien reichte. Dieses Ziel wurde trotz des Hinzutretens der Triangulation Kurhessens durch den Gaußschüler Gerling nicht erreicht. Nicht gut zu machen war auch der Umstand, daß die Basismessung Schumachers nie gehörig reduziert wurde. Gauss legte seine Dreieckspunkte in der Hauptsache an die gleichen Stellen, welche bei der französischen Kriegsvermessung Hannovers von den französischen Vermessungsoffizieren ausgewählt waren. Dies geschah wohl deshalb, weil er selbst bei Beginn der Messung in diesen Dingen noch keine Erfahrung besaß und er deshalb aus der größeren praktischen Erfahrung seiner Vorgänger Nutzen ziehen wollte. Viele Vorwürfe, die sich gegen die Auswahl der Dreieckspunkte und andere Umstände richten, sind sicher gegenstandslos, da sie Verhältnisse als gegeben voraussetzen, die damals nicht vorlagen, und da sie keine Rücksicht auf die Knappheit der Geldmittel nehmen, mit denen Gauss auszukommen hatte. So hat schließlich das ganze Unternehmen nicht zum erhofften Ergebnis geführt. Dennoch ist es in der Geschichte

des Vermessungswesens von einzigartiger Bedeutung durch die vielen Untersuchungen und Erfindungen, die mit ihm zusammenhängen.

Als erstes ist die Erfindung des Heliotropen zu nennen, durch den eine anzuzielende Station markiert wird durch das auf ihr in einem Spiegel nach dem Beobachter reflektierte Sonnenbild. Dadurch wurde zum erstenmal die Winkelmessung in Dreiecken mit sehr langen Seiten mit der nötigen Sicherheit möglich. Die ursprüngliche Konstruktion von GAUSS war ein Meisterstück der Erfindungsgabe. Man sah bald, daß sie durch wesentlich einfachere Einrichtungen ersetzbar war. Neben dem unmittelbaren Gebrauch in der Vermessung benutzte GAUSS den Heliotropen auch zur Übermittlung vereinbarter Botschaften zwischen den Eckpunkten der Dreiecke durch Lichtblitze. Er wies auch nachdrücklich auf diese Möglichkeit der Lichttelegraphie hin, die ihm selbst bei den elenden Post- und Verkehrsverhältnissen seiner Zeit große Zeitverluste ersparte. Es darf wohl daran erinnert werden, daß diese Art der Telegraphie in Kolonialkriegen kurz nach 1900 eine bedeutende Rolle spielte.

GAUSS trug nicht nur die Hauptlast der Messungen, sondern er mußte auch alle mit den Messungen zusammenhängenden Rechnungen selbst durchführen. Die persönliche intime Kenntnis aller Einzelheiten der Messungen und der Rechnungsführung ermöglichte es ihm, für diese Messungen und Rechnungen vorbildliche Anweisungen zù geben, die ihn zum eigentlichen Vater der modernen Geodäsie gemacht haben. Durch ihn ist die Ausgleichsrechnung zum Hauptwerkzeug des rechnenden Geodäten geworden. Aus persönlicher Erfahrung stammt auch die Erkenntnis der mangelnden Genauigkeit der Winkelmessung durch Repetition und die Erkenntnis der Häufigkeit der Lotabweichungen (d.h. des Nichtzusammenfallens der Normale des der Vermessung als Bezugsfläche zu Grunde liegenden Ellipsoids mit der Normalen auf der Niveaufläche der Schwerkraft). Über das Zustandekommen der Lotabweichungen macht sich dabei GAUSS völlig zutreffende Vorstellungen.

Weiter führen Fragen der Geodäsie GAUSS zu seinen Untersuchungen über konforme Abbildung und Differentialgeometrie. Die von GAUSS seit 1795 benutzte Ausgleichsrechnung wurde unbhängig auch von LEGENDRE entdeckt, der sie « Methode der kleinsten Quadrate » nannte. Die erste Darstellung durch GAUSS enthält die Theoria motus, durch die Verzögerungen im Druck dieses Werkes kommt aber die Priorität der Veröffentlichung LEGENDRE zu. Die Ausgleichsrechnung hatte zunächst den Zweck der im 18. Jahrhundert sehr verbreiteten Gewohnheit entgegen zu treten, das Ergebnis von Meßreihen dadurch zu verfälschen, daß man auf Grund irgend welcher Voreingenommenheiten einen Teil der Messungen verwarf und nur die beibehielt, welche das erwünschte Ergebnis lieferten. GAUSS steht auf dem Standpunkt, daß grund-

sätzlich alle Messungen einer Meßreihe gleichberechtigt sind und daß die Grundsätze, nach denen eine Meßreihe zu behandeln ist, von vornherein feststehen müßen. Es handelt sich also um eine Art Gesetzgebung. Einleuchtende praktische Erwägungen führen leicht zur Aufstellung von Regeln einschließlich solcher, die zur Beurteilung der Genauigkeit des Ergebnisses dienen. Auf diesem Standpunkt blieb LEGENDRE stehen. GAUSS suchte aber den Verfahrensregeln eine tiefere Grundlage zu geben, indem er sie mit der Wahrscheinlichkeitsrechnung in Verbindung brachte. Da es grundsätzlich unmöglich ist, die wahren Werte der Größen zu bestimmen, welche aus den Messungen ermittelt werden sollen, setzt sich GAUSS das Ziel, die wahrscheinlichsten Werte für sie festzustellen. Dazu braucht er aber die Fehlerfunktion $\varphi(\varepsilon)$, d.h. die Funktion, mit deren Hilfe man bei beliebig vorgegebenen a und b $(a < b)$ die Wahrscheinlichkeit dafür daß der Fehler ε zwischen a und b liegt in der Form schreiben kann $\int_a^b \varphi(\varepsilon)\,\mathrm{d}\varepsilon$. Die Bestimmung dieses $\varphi(\varepsilon)$ ist nicht ohne ein Axiom möglich. So beruft sich GAUSS auf den allgemein anerkannten Grundsatz, daß bei direkten Messungen einer Größe das arithmetische Mittel der Messungen für sie einen brauchbaren Wert ergibt. Dies führt dann zum sog. « Gaußschen Fehlergesetz » $(h/\sqrt{\pi})\exp[-h^2\varepsilon^2]$. Dabei ist h eine Konstante, die man als Maß für die Genauigkeit der Beobachtungen ansehen kann.

Mit dieser Begründungsart der Ausgleichsrechnung war GAUSS aus verschiedenen Gründen nicht zufrieden. Vor allem war er sich des hypothetischen Charakters seines Fehlergesetzes vollkommen bewußt. Auch fand er, daß die Auszeichnung des wahrscheinlichsten Wertes der Natur der Aufgabe nicht gerecht werde, da dem wahrscheinlichsten Wert nur eine « unendlich kleine » Wahrscheinlichkeit zukommt. Er gab deshalb eine Neubegründung in der dreiteiligen Abhandlung « Theoria combinationis observationum erroribus minimis obnoxiae » (Werke IV) von 1821/1823/1826. Der Grundgedanke ist hier der Vergleich der Aufgabe, irgend welche Größen durch Messungen zu bestimmen, mit einem Glücksspiel, bei dem man nur verlieren und nie gewinnen kann. Handelt es sich um ein echtes Spiel, so muß der Spieler noch gewisse Verfügungsmöglichkeiten haben. Diese wird er zweckmäßig dazu benutzen, das Risiko seines Spieles möglichst niedrig zu halten. In unserem Fall kann man zunächst noch frei verfügen über die Größen, die man durch Messungen bestimmen will. Hat man sie gewählt, so folgen daraus die Werte für die zu messenden Größen. Diese Werte werden von den Meßergebnissen abweichen, Die Differenzen zwischen Meßergebnis und berechnetem Wert sind die Fehler. Jedem Fehler ε antspricht ein « Verlust » $V(\varepsilon)$, das Risiko ist dann die Summe aller Produkte $V(\varepsilon)$ mal der Häufigkeit von ε. Die zweckmäßigste Wahl der zu bestimmenden Größen ist die, welche dieses « Risiko » zu einem Minimum

macht. Bestimmt werden sie aber erst durch die vorgeschriebene Spielregel, d.h. durch Verfügung über die Funktion $V(\varepsilon)$. Das einfachste und für die Rechnung bequemste ist hier $V(\varepsilon) = \varepsilon^2$ zu nehmen. Dieser Ansatz führt dann ohne spezielle Annahmen über Wahrscheinlichkeiten zum Ziel und zwar nicht nur bei direkten, sondern auch bei vermittelnden und bedingten Beobachtungen. GAUSS hat sich nun nicht nur mit der allgemeinen Theorie beschäftigt, sondern das Rechenverfahren, das man in jedem konkreten Fall anzuwenden hat, auf das feinste durchgearbeitet. Es ist auch heute noch im Gebrauch, soweit nicht moderne Rechenhilfsmittel Abänderungen erforderten. Die Anwendungen, die GAUSS machte, liegen in der Astronomie, z.B. bei der Verbesserung einer berechneten Planetenbahn beim Vorliegen neuer Beobachtungen (« Disquisitio de elementis ellipticis Palladis », Werke VI) und bei der « Bestimmung des Breitenunterschiedes zwischen den Sternwarten von Göttingen und Altona » (Werke IX), sowie in der Geodäsie (Schluß der Theoria combinationis). In den Allgemeingebrauch eingeführt wurde die Ausgleichsrechnung, da den Praktikern die Abhandlungen von GAUSS zu schwer waren, erst durch die Gaußschüler ENCKE, HAGEN und GERLING.

An die Gradmessung in Hannover schloß sich eine Triangulation des Landes an. Zur Berechnung der Dreiecke benutzte GAUSS eine konforme (winkeltreue) Abbildung des Ellipsoids in die Ebene. Er konnte sich damit auf Berechnungen in der Ebene beschränken und für alle Punkte rechtwinklige ebene Koordinaten berechnen. x-Achse seines rechtwinkligen Koordinatensystems war der Meridian der Sternwarte in Göttingen, der Koordinatenanfang lag in Göttingen. Veröffentlicht hat GAUSS über diese Projektionsmethode nichts. Erst durch L. KRÜGER wurde sie der Vergessenheit entrissen, fortgebildet und dem deutschen Vermessungswesen zu Grunde gelegt. Vor dieser direkten Abbildung des Erdsphäroids in die Ebene benutzte die Preussische Landesaufnahme für besondere Zwecke die ebenfalls auf Entwicklungen von GAUSS zurückgehende Gauss-Schreibersche konforme Doppelprojektion. Bei dieser wird zunächst das Erdsphäroid konform auf die Kugel und dann diese konform in die Ebene abgebildet. Mit seiner konformen Projektion schuf GAUSS etwas wesentlich Neues, da bis dahin allgemein die sog. Cassini-Soldnerschen Koordinaten benutzt wurden.

Die theoretischen Grundlagen gab GAUSS in den Arbeiten: « Allgemeine Lösung der Aufgabe: Die Theile einer gegebenen Fläche auf einer anderen Fläche so abzubilden, daß die Abbildung dem Abgebildeten in den kleinsten Theilen ähnlich wird » (Werke IV) von 1822 und « Untersuchungen über Gegenstände der höheren Geodäsie I/II » (Werke IV) von 1843/1846. Die erste Abhandlung ist die Bearbeitung einer auf Vorschlag von SCHUMACHER gestellten, aber auf GAUSS zurückgehenden Preisaufgabe der Kopenhagener Gesellschaft

der Wissenschaften. Gauss zeigt in ihr: Ist das Quadrat des Linienelements einer beliebigen reellen Fläche in der Gestalt $ds^2 = E\,dp^2 + 2F\,dp\,dq + G\,dq^2$ gegeben, so kann man stets eine Parametertransformation so angeben, daß ds^2 die Gestelt annimmt $ds^2 = \psi(\lambda, \mu)(d\lambda^2 + d\mu^2)$. Er zerlegt dazu ds^2 in zwei konjugiert komplexe lineare Differentialformen in dp und dq und integriert die Differentialgleichung $ds^2 = 0$ mittels eines Eulerschen Multiplikators für jeden der Faktoren. Mit der neuen Form des Linienelements ist aber die konforme Abbildung der gegebenen Fläche in die Ebene gegeben. Damit ist auch die konforme Abbildung von Flächen aufeinander erledigt. In der späteren Abhandlung von 1843/1846 behandelt Gauss den Sonderfall der konformen Abbildung des Ellipsoids auf die Kugel.

Auch die Untersuchungen von Gauss zur Differentialgeometrie hängen auf auf das engste mit seinen geodätischen Arbeiten zusammen. Nimmt man nicht auch den Raum außerhalb der Fläche hinzu, so kann man auf ihr nur Größen von Winkeln und Längen von Kurven messen. Diese aber ändern sich nicht, wenn man die Fläche ohne Dehnung verbiegt. Insbesondere bleiben dabei geodätische (d.h. kürzeste) Linien stets kürzeste. Während die Differentialgeometrie vor Gauss nur Eigenschaften der Flächen untersuchte, die bei Bewegungen der Fläche ohne Formänderung erhalten bleiben, mußten sich für Gauss von seinem Ausgangspunkt her als wesentlich neu die biegungsinvarianten Eigenschaften hinzugesellen.

Die die Untersuchungen von mehr als einem Jahrzehnt zusammenfassende abschließende Veröffentlichung « Disquisitiones generales circa superficies curvas » (Werke IV) von 1828 stellt die systematische Darstellung voran und weicht so ganz erheblich von der an den Entwürfen zu verfolgenden geschichtlichen Entwicklung ab. Gauss betrachtet im Anschluß an Euler die Koordinaten der Flächenpunkte als Funktionen zweier Parameter p und q. Er beschränkt sich auf die Betrachtung regulärer Flächenpunkte, in denen man eindeutig eine Normale erklären kann. Diese wird mit einer Richtung versehen. Jedem Flächenstück ordnet er ein « sphärisches Bild » zu, indem er jedem Punkt des Flächenstückes den Punkt der Einheitskugel entsprechen läßt, der auf dieser durch den der Flächennormale parallelen und gleichgerichteten Radius ausgeschnitten wird. Dieses sphärische Bild ist offenbar von der Gewohnheit der Astronomen hergenommen, Richtungen durch Punkte der Himmelskugel festzulegen. Jedem Flächenstück schreibt Gauss als « Gesamtkrümmung » (curvatura integra) den orientierten d.h. mit Vorzeichen versehenen Flächeninhalt des sphärischen Bildes zu. Bildet man das Verhältnis der Gesamtkrümmung eines Flächenstücks zum Inhalt des Flächenstücks und läßt dann das Flächenstück sich auf einen Punkt P zusammenziehen, so nennt man den Limes, dem dabei das genannte Verhältnis zustrebt, das Krümmungsmaß k der Fläche

im Punkt P. Aus dieser Erklärung folgt sofort, daß die Gesamtkrümmung $\int k\,d\sigma$ geschrieben werden kann, wobei die Integration über das ganze Flächenstück (Flächenelement $d\sigma$) zu erstrecken ist. Nach Herleitung der Darstellung $k = 1/R_1R_2$ (R_1R_2 die Hauptkrümmungsradien) stellt GAUSS k in der Gestalt $k = (DD'' - D'^2)/(EG - F^2)$ dar. Dabei sind D, D', D'' nichts anderes als die heute meist mit L, M, N bezeichneten Koeffizienten der zweiten Fundamentalform und E, F, G die Koeffizienten der ersten Fundamentalform, d.h. des Quadrats des Linienelementes $ds^2 = E\,dp^2 + 2F\,dp\,dq + G\,dq^2$. Dies wird erst nachträglich gezeigt. Aus dieser Darstellung von k folgert GAUSS, daß k allein durch E, F, G und ihre partiellen Ableitungen erster und zweiter Ordnung nach p und q ausdrückbar ist. Da ds biegungsinvariant ist, hat er damit den von ihm « theorema egregium » genannten Satz erhalten: Wenn eine krumme Fläche auf eine andere abgewickelt werden kann, so bleibt damit das Krümmungsmaß in den einzelnen Punkten ungeändert. Analoges ergibt sich daraus für die Gesamtkrümmung.

GAUSS wendet sich dann Untersuchungen über geodätische Linien zu. Ihre Differentialgleichung wird aufgestellt und der Satz hergeleitet: Hat man auf einer Fläche ein Feld von geodätischen Linien (eingeschlossen den Fall, daß alle diese Geodätischen durch einen Punkt gehen), so schneiden zwei beliebige Orthogonaltrajektorien dieses Feldes auf allen geodätischen Linien des Feldes gleiche Bogen aus. Anschließend betrachtet GAUSS auf der Fläche ein Dreieck, dessen Seiten geodätische Linien sind. Die Winkel in den Ecken seien A, B, C. Er zeigt dann, daß die Gesamtkrümmung dieses Dreiecks gleich $A + B + C - \pi$ ist. Dies ist eine weitgehende Verallgemeinerung eines bekannten sehr elementaren Satzes der sphärischen Trigonometrie. – LEGENDRE hatte die Frage gestellt, welche Beziehungen zwischen den Winkeln A, B, C eines sphärischen Dreiecks und den Winkeln A^*, B^*, C^* des ebenen Dreiecks mit denselben Seitenlängen bestehen. Er hatte gefunden, daß für kleine sphärische Dreiecke in guter Näherung ist $A - A^* = B - B^* = C - C^* = \frac{1}{3}(A + B + C - \pi)$. Diesen Satz kann jetzt GAUSS verallgemeinern auf von geodätischen Linien gebildete kleine Dreiecke auf beliebigen Flächen. Die Differenzen $A - A^*$, $B - B^*$, $C - C^*$ sind dann nicht gleich, sondern hängen vom Krümmungsmaß der Fläche in den Ecken des Dreiecks ab. GAUSS bemerkt aber, daß für Flächen, die wenig von der Kugel abweichen, der Unterschied der von ihm angegebenen Werte dieser Differenzen gegen den Legendreschen Wert fast unmerklich ist. Er belegt das durch das Beispiel des größten von ihm gemessenen Dreiecks mit den Eckpunkten Hohehagen, Brocken und Inselsberg (Seitenlängen etwa 69 km, 85 km, 107 km).

Wir müßen es uns versagen, auf die späteren, nur im Nachlaß erhaltenen Untersuchungen von GAUSS zur Differentialgeometrie einzugehen. Ebenso

übergehen wir die aus physikalischen Fragestellungen entsprungenen topologischen Untersuchungen. Nur auf eines sei hingewiesen: GAUSS hat Räume beliebiger Dimensionszahl in seine Betrachtungen einbezogen. Er steht damit allein unter allen Mathematikern seiner Zeit.

* * *

Die ausgedehnten Beschäftigungen von GAUSS mit physikalischen Fragen haben als eigentlichen Mittelpunkt seine theoretisch-physikalischen Untersuchungen. Wir beginnen mit der kurzen Besprechung des von ihm entdeckten Prinzips des kleinsten Zwanges, um dann zu den wichtigen Untersuchungen zur Potentialtheorie überzugehen.

LAGRANGE hatte in seiner Mécanique analytique die Statik auf das Prinzip der virtuellen Verrückungen, die Dynamik auf das d'Alembertsche Prinzip gegründet. GAUSS gibt nun in seiner Abhandlung « Über ein neues allgemeines Grundgesetz der Mechanik » (Werke V) von 1829 das sog. « Prinzip des kleinsten Zwanges » an, das Statik und Dynamik zugleich umfaßt. Er begründet es, indem er zeigt, daß unter den für das d'Alembertsche Prinzip geltenden Voraussetzungen dieses aus seinem neuen Prinzip folgt. Das bedeutet natürlich nicht die Gleichwertigkeit der beiden Grundsätze. Wir wissen heute, daß das Gaußsche Prinzip viel weiter reicht und wohl das umfassendste aller mechanischen Prinzipien ist. Damit entfällt aber die Frage nach der Beweisbarkeit, es ist eben das Grundaxiom der Mechanik und könnte als solches nur aus gleichwertigen Axiomen deduziert werden. Die von GAUSS gegebene Begründung kann nur darlegen, wie man folgerichtig auf das neue Prinzip kommt. Gegeben seien n Massenpunkte (x_i, y_i, z_i) mit den Massen m_i. Auf diese mögen die Kräfte (X_i, Y_i, Z_i) wirken. Bedeuten, wie üblich, Punkte Ableitungen nach der Zeit, so nennt GAUSS Zwang den Ausdruck

$$Z = \Sigma \frac{1}{m_i}\left\{(X_i - m_i\ddot{x})^2 + (Y_i - m_i\ddot{y}_i)^2 + (Z_i - m_i\ddot{z}_i)^2\right\}.$$

GAUSS geht nun von einem gegebenen Zustand des Punktsystems, d.h. von gegebenen (x_i, y_i, z_i) und $(\dot{x}_i, \dot{y}_i, \dot{z}_i)$ aus und fordert, daß die $(\ddot{x}_i, \ddot{y}_i, \ddot{z}_i)$ so gewählt werden, dass Z unter Berücksichtigung der Nebenbedingungen (einschließlich ihrer Ableitungen nach der Zeit) ein Minimum wird. Die Nebenbedingungen können Gleichungen und Ungleichungen sein. Es wird nicht gefordert, wie beim d'Alembertschen Prinzip, daß sie in den Differentialen der Koordinaten linear sind. Das Gaußsche Prinzip ist also, anders als die gewohnten Prinzipe, kein Variationsprinzip, sondern ein Minimalprinzip. Formal ordnet es sich den Formeln der Ausgleichrechnung unter. GAUSS bemerkt, daß es merkwürdig sei, daß die Natur die freien Bewegungen eines mechanischen

Systems unter dem Druck von Bedingungen ebenso modifiziert, wie der rechnende Mathematiker Erfahrungen über unter einander abhängige Größen ausgleicht. Denn letzterer verfährt ja nach einem « Prinzip der Gerechtigkeit ».

1830 veröffentlichte GAUSS seine « Principia generalia theoriae fluidorum in statu aequilibrii » (Werke V). Diese Arbeit befaßt sich mit der Theorie der Kapillarkräfte. Zu Grunde liegt die Annahme, daß die molekularen Kräfte Zentralkräfte der Gestalt $m_1 m_2 f(r_{12})$ sind, wobei m_1 und m_2 die Massen der Partikeln und r_{12} ihre Entfernung ist. Dabei soll $f(r)$ mit wachsendem r außerordentlich rasch gegen Null abnehmen. Ausgehend von endlich vielen Partikeln und nachfolgendem Grenzübergang zum Kontinuum, liefert das Prinzip der virtuellen Verrückungen die Aussage, daß im Gleichgewicht das Potential der wirkenden Kräfte einen stationären Wert haben muß. Sind als Kräfte zugelassen die Schwere, die molekulare Anziehung der Flüssigkeitsteilchen aufeinander und die Anziehung der Gefäßwand auf die Flüssigkeitsteilchen, so zerfällt das zu bestimmende Potential in verschiedene einzeln bestimmbare Teile. Macht man noch Gebrauch davon, daß für « endliche » Entfernungen die Molekularkräfte Null sind, so erhält man für das Potential K (die hier benutzten Bezeichnungen weichen von GAUSS ab) den Ausdruck

$$-K = \varepsilon g \int z \, \mathrm{d}v + \alpha^2 F + (\alpha^2 - \beta^2) G + \gamma^2 V .$$

Dabei bedeuten ε die Dichte der Flüßigkeit, g die Schwerebeschleunigung, α, β, γ experimentell zu bestimmende Konstanten, F die freie Oberfläche, G die von der Flüßigkeit benetzte Fläche der Wand, V das Volumen der Flüßigkeit, $\mathrm{d}v$ das Volumelement und z die vertikale Koordinate. Die weiteren Ausführungen dieser Abhandlung gehören in die Variationsrechnung für Doppelintegrale. Es sei für sie auf die ausführliche Darlegung von BOLZA (Werke X_2) verwiesen.

In die Potentialtheorie gehört im Grunde schon die Arbeit über die Anziehung eines elliptischen Ringes, sodann die Abhandlung « Theoria attractionis corporum sphaeroidicorum ellipticorum homogeneorum methodo nova tractata » (Werke V) von 1813. In beiden Arbeiten tritt der Begriff des Potentials noch nicht auf. Die Anziehung homogener Ellipsoide war ein altes, aber erst durch LAPLACE und IVORY gelöstes Problem. Beide benutzen unendliche Reihen in der der Zeit entsprechenden bedenkenlosen Weise. GAUSS geht hier ganz andere Wege. Er beginnt mit sechs Sätzen über Integrale über eine beliebige geschlossene Fläche S. Die meisten sind im Grunde Verwandlungen von Raumintegralen über das Innere von S in Oberflächenintegrale, ein uns heute ja vertrautes Verfahren. Wegen seiner späteren Verwendung ist besonders bemerkenswert das Integral, welches die « scheinbare Größe von S gesehen von einem Punkt M aus » (den « räumlichen Winkel », unter dem S

von M aus erscheint) angibt. Dieses ist, je nachdem ob M außerhalb, auf oder innerhalb S liegt, gleich 0, oder 2π oder 4π. Es ist zugleich das einfachste Doppelbelegungspotential. Aus den sechs Integralsätzen folgt dann die Lösung des Ellipsoidproblems.

Das Potential wird von GAUSS erst eingeführt in der grundlegenden Abhandlung « Allgemeine Lehrsätze in Beziehung auf die im umgekehrten Verhältnisse des Quadrats der Entfernung wirkenden Anziehungs- und Abstoßungskräfte » (Werke V) von 1839. Sätze aus dieser Abhandlung teilte GAUSS schon früher mit in seinen Arbeiten zum Magnetismus. Er macht dort von ihnen wichtige Anwendungen. Der Abhandlung von GAUSS geht zeitlich voran die Arbeit von G. GREEN « Essay on the application of mathematical analysis on the theories of electricity and magnetism » von 1828. Die Ergebnisse dieser Arbeit überschneiden sich mit denen der Arbeit von GAUSS. Es steht fest, daß GAUSS die Arbeit von GREEN nicht gekannt hat und daß die Verbreitung des Potentialbegriffes allein ihm zu danken ist. Der Name « Potential », der GAUSS und GREEN gemeinsam ist, tritt schon im 18. Jahrundert auf und entstammt der philosophischen Terminologie.

GAUSS beginnt mit der Herleitung der Eigenschaften von Potentialen räumlicher und flächenhafter Massenverteilungen (Stetigkeits- und Unstetigkeitsverhalten der Potentiale und ihrer partiellen Ableitungen erster und zweiter Ordnung, Erfüllung der Laplaceschen Differentialgleichung im massefreien Raum, der Laplace-Poissonschen innerhalb der anziehenden Massen), die nach ihm « in ihrem wesentlichen Inhalt nicht neu » seien. Ist V das Potential von Massen M_j, die in den Punkten P_j konzentriert sind, ebenso v das Potential von Massen m_k in den Punkten p_k, so ist $\sum M_j v(P_j) = \sum m_k V(p_k)$. Durch Grenzübergang folgert GAUSS daraus den entsprechenden Satz für zwei Potentiale irgendwelcher (räumlicher, flächenhafter oder punktförmiger) Massenverteilungen. Ein wichtiger Spezialfall ist der Satz: Gegeben das Potential V irgendwelcher Massen und eine Kugel vom Radius R. Dann ist der Mittelwert von V auf der Kugel $(1/4\pi R^2)\int V \mathrm{d}S = V_0 + M_0/R$, wobei V_0 der Wert von V im Kugelmittelpunkt ist und M_0 der Teil der V erzeugenden Massen, der in der Kugel liegt. Untersuchungen von Integralen über eine geschlossene Oberfläche S (innere Normale n) schließen sich an, wie z.B. $\int(\partial V/\partial n)\,\mathrm{d}S$ und $\int V(\partial V/\partial n)\,\mathrm{d}S$ (V irgend ein Potential). Hinzu treten Sätze über Potentiale, die in Stücken des Raums oder auf geschlossenen Flächen konstant sind. Ist insbesondere ein Potential V auf einer geschlossenen Fläche S konstant und liegen alle Massen, zu denen V gehört, im Innern (Äußeren) von S, so ist V außerhalb (innerhalb) von S konstant. GAUSS betrachtet weiter eine geschlossene Fläche S und belegt diese mit Masse von überall positiver Dichte m und der Gesamtmasse M. Es sei V das Potential dieser Belegung und U eine

stetige Funktion auf S. Bildet man für alle möglichen zulässigen Wahlen der Belegung m das Integral $\int(V-2U)m\,dS$, so ist die Menge dieser Integrale nach unten beschränkt. Daraus schließt GAUSS (und das ist im Grunde derselbe anfechtbare Schluß, der beim Dirichletschen Prinzip benutzt wird), daß es eine spezielle Massenbelegung m gibt, die das Integral zu einem Minimum macht. Für diese Belegung ist dann $V-U=\text{const}$ überall da, wo $m>0$ ist. Ist speziell $U=0$ auf ganz S, so kann man zeigen, daß die zugehörige Minimalbelegung auf ganz S positiv ist. Das zugehörige Potential ist auf S konstant. Aus diesen beiden ausgezeichneten Belegungen läßt sich nun eine dritte herstellen, derart, daß für ihr Potential V^* gilt $V^*-U=0$ auf ganz S. Das bedeutet: Ist das U auf S vorgegeben, so gibt es stets ein Potential, das auf S diese Werte annimmt. Sind insbesondere U die Werte, welche das Potential irgendwelcher ganz innerhalb S (außerhalb S) gelegener Massen auf S annimmt, so gibt es eine Flächenbelegung von S, deren Potential außerhalb S (innerhalb S) mit dem Potential der gegebenen Massen übereinstimmt. Der Nachlaß von GAUSS zeigt, daß er diese Untersuchungen weitergeführt und sich auch mit dem logarithmischen Potential beschäftigt hat. Dessen Zusammenhänge mit der Funktionentheorie und der konformen Abbildung waren ihm bekannt.

Als 1829 der belgische Physiker QUETELET im Garten der Göttinger Sternwarte magnetische Beobachtungen machte, wies GAUSS ihn darauf hin, daß man die Schwingungsdauer eines Magneten aus den Zeiten der größten Elongationen nur sehr ungenau ermitteln könne. Empfehlenswerter sei, Beobachtungen in der Nähe der Ruhelage zu verwenden. Dieses Prinzip, zur genauen Messung von Zeiten eine bewegte Marke dort zu beobachten, wo ihre Geschwindigkeit am größten ist, zur genauen Feststellung des Ortes aber zu den Zeiten kleinster Geschwindigkeit zu messen, ist heute allen geläufig. In allgemeine Aufnahme wurde es gebracht durch GAUSS' « Anleitung zur Bestimmung der Schwingungsdauer einer Magnetnadel » (Werke V) von 1837. Diese Arbeit, die sich mit schwach gedämpften Schwingungen befaßt, ist wohl eine der ersten, welche die Theorie zur zweckmäßigen Einrichtung des Beobachtungsverfahrens nutzbar zu machen versucht. Die Regeln für die Bestimmung der Gleichgewichtslage und zur Bestimmung der Dämpfung mittels des logarithmischen Dekrements stammen aus ihr. Zu diesen Regeln treten noch die anzubringenden Korrekturen: Ermittlung der Schwingungsdauer der ungedämpften aus der Dauer der gedämpften Schwingung, Veränderung der Gleichgewichtslage durch diese Reduktion, Übergang zu unendlich kleinen Schwingungen, Einfluß der Torsion des Aufhängefadens u.s.w.. Nichtberücksichtigt ist bei GAUSS lediglich die Reduktion auf den luftleeren Raum.

Das Interesse von GAUSS an Fragen des Magnetismus und der Elektrizitätslehre fiel in eine Zeit stürmischer Entwicklung dieser Disziplinen: 1820 Ent-

deckung des Elektromagnetismus durch Oerstedt, 1820/21 Biot-Savartsches Gesetz, 1822/24 grundlegende Arbeiten Ampères, 1827 Ohmsches Gezetz, 1831 Entdeckung des Elektromagnetismus durch Faraday. Schon 1831 wurde die magnetische Deklination in Göttingen regelmäßig beobachtet und ihre Störung durch ein Nordlicht bemerkt. Entscheidend war aber das Interesse, das Alexander v. Humboldt als Geograph und Geophysiker den Erscheinungen des Erdmagnetismus entgegenbrachte. Es gelang ihm, die Mitwirkung von Gauss für diese Forschungen zu erlangen. Dabei handelte es sich zunächst um die Sammlung eines hinreichend großen Beobachtungsmaterials, die Ausgestaltung der Apparate und Beobachtungsmethoden und die zweckmäßige Einrichtung und Organisation der Beobachtungen. Die Führung lag bald ganz allein bei Gauss und seinem Mitarbeiter, dem 1831 nach Göttingen berufenen Physiker W. Weber. Zur Unterrichtung der Mitarbeiter und zur Verbreitung der Ergebnisse ihrer Forschungen ließen Gauss und Weber auf ihre Kosten eine Zeitschrift « Resultate aus den Beobachtungen des magnetischen Vereins » erscheinen, die es auf sechs Bände brachte (1836-1841). Der « Magnetische Verein » beruhte auf völlig freier Zusammenarbeit: es handelte sich also um eines der frühesten Beispiele dessen, was man heute « teamwork » nennt. Es wurden Beobachtungen nicht nur aus Europa, sondern von der ganzen Erde, vom Nordkap bis zum Kap der guten Hoffnung, aus Nord- und Südamerika, aus Sibirien, aus Süd- und Ostasien und Australien bis nach Tahiti hin zusammengetragen. Beobachtet wurden zunächst Deklination, Horizontalintensität und Inklination. Die Beobachtungen der Inklination traten bald zurück, da sie sich nicht mit der gleichen Genauigkeit messen ließ, wie Deklination und Horizontalintensität und da die potentialtheoretischen Untersuchungen von Gauss erwiesen, daß man auf die Messung der Inklination verzichten konnte, allerdings unter der Annahme, daß der Erdmagnetismus seinen Sitz völlig im Innern der Erde hat. Es sei vorausgenommen, daß Gauss die damit gestellte Randwertaufgabe auch zahlenmäßig gelöst hat. Er hat die Ergebnisse seiner Rechnungen in dem « Atlas des Erdmagnetismus » (Werke XII) von 1840 unter verschiedenen Gesichtspunkten kartenmäßig dargestellt und die errechneten und beobachteten Werte für Deklination, Inklination und Horizontalintensität tafelmäßig verzeichnet und verglichen. Die Lösung der Randwertaufgabe beruht hier auf einer Entwicklung des Potentials in eine Reihe nach allgemeinen (Laplaceschen) Kugelfunktionen, von der Gauss die (übrigens physikalisch deutbaren) Glieder bis zur vierten Ordnung beibehält. Weiter zu gehen erlaubte die Genauigkeit der Beobachtungen nicht. Für die mitgenommenen Glieder werden die Genauigkeitsmängel der Beobachtungen kompensiert durch ihre verhältnismäßig große Zahl. Diese Rechnung ergab, daß die Erde als Magnet genau zwei Pole besitzt. Es war ein Triumph der Rechnung, daß diese Pole von Forschungsreisenden in nicht allzugroßer Entfernung vom berechneten Ort wirklich aufgefunden wurden.

Das anschauliche Bild, mit dem GAUSS bei seinen Überlegungen über den Magnetismus arbeitet, ist das zweier magnetischer Flüssigkeiten. Bei dieser Vorstellung muß noch erklärt werden, was ein magnetischer Pol ist. Dazu ersetzt GAUSS die räumlich verteilten magnetischen Flüssigkeiten durch die im Außenraum gleichwertige, d.h. dasselbe Potential liefernde, Flächenbelegung einer alle magnetischen Flüssigkeiten einschließenden geschlossenen Fläche. In der « Intensitas vis magneticae terrestris ad mensuram absolutam revocata » (Werke V) von 1832 behandelt GAUSS die Festsetzungen, die nötig sind, um an verschiedenen Orten oder zu verschiedenen Zeiten durchgeführte magnetische Messungen vergleichbar zu machen. Dabei schließt er sich an das französische metrische System an, wohl weil dieses damals das einzige war, das in einem ausgedehnten Gebiet Geltung besaß und durch Urmaße hinreichend gesichert war. Die Quantität des Magnetismus bestimmt er durch die Kraftwirkung unter Zugrundelegung der von COULOMB festgestellten und von ihm noch nachgeprüften Tatsache, daß für die magnetische Anziehung und Abstoßung formal das Newtonsche Attraktionsgesetz gilt. Die magnetische Einheit übt dann in der Entfernung 1 mm auf die Masse 1 mg die Kraft 1 $mg \cdot mm \cdot s^{-2}$ aus. Der bei dem Anziehungsgesetz noch mögliche, der Gravitationskonstanten entsprechende Faktor wird zu 1 normiert. Die Art und Weise, wie Deklination und Horizontalintensität nach den Vorschriften von GAUSS bestimmt werden, ist heute Allgemeinbesitz und braucht wohl nicht geschildert zu werden. Es sei auf einige Umstände aufmerksam gemacht. Da GAUSS Präzisionsmessungen anstrebt, benutzt er die Ablesung einer Skala mit Spiegel und Fernrohr, die er unabhängig von POGGENDORFF erfunden hatte. Ferner bringt er bei Schwingungsbeobachtungen stets die früher erwähnten Korrekturen an. Wegen der thermischen Abhängigkeit des Magnetismus sorgt er für möglichste Konstanz der Temperatur. Die Beobachtungen der Deklination an verschiedenen Beobachtungsorten ergaben, daß die bedeutenden Schwankungen der Deklination in kurzen Zeiten keine Störungen lokaler Natur waren. Sie traten bei allen Beobachtungsorten in nahezu der gleichen Größe und zur gleichen Zeit auf.

GAUSS vermutete, daß entsprechendes auch für die Horizontalintensität gelten müsse. Bei dem üblichen Meßverfahren für diese ließ sich das nicht feststellen, da wegen der verhältnismäßig langen Dauer der Messung diese nur einen Mittelwert ergeben kann. In dem Bifilarmagnetometer schuf sich GAUSS ein Instrument zur Messung rascher Schwankungen der Horizontalintensität. Dabei ist dieses Instrument auch zu anderen Messungen brauchbar. Trotz seiner Erkenntnis, daß die Messung der Inklination nicht mit der gleichen Schärfe möglich ist, wie die der beiden anderen Elemente des Erdmagnetismus und trotz ihrer geringeren Wichtigkeit hat sich GAUSS auch um ihre Bestimmung bemüht. Als geeignetstes Instrument findet er den Erdinduktor.

Eine wesentliche Hilfe bei seinen Untersuchungen war für GAUSS der Bau eines eisenfreien magnetischen Observatoriums im Jahr 1833. In seiner Arbeit « Allgemeine Theorie des Erdmagnetismus » (Werke V) von 1838 gibt GAUSS die Zusammenstellung der potentialtheoretischen und anderen Sätze, die wir im Verlauf unserer Darstellung schon erwähnt haben, so daß wir auf sie nicht weiter einzugehen brauchen. Manche Einzelheit ist allerdings bemerkenswert. So benutzt er z.B. den Satz, daß die Arbeit auf einem geschlossenen Weg im Potentialfeld Null ist, zu folgender Berechnung, bei der die auftretenden Integrale numerisch angenähert werden. Er betrachtet das erdmagnetische Feld und berechnet die Horizontalintensität in Paris aus der Deklination in Paris und den Deklinationen und Horizontalintensitäten in Göttingen und Mailand. Integrationsweg sind die Seiten des Dreiecks Göttingen-Mailand-Paris. Die Übereinstimmung des errechneten Wertes mit dem beobachteten ist überraschend gut.

Schon bald nach Beginn seiner Untersuchungen zum Magnetismus stellte GAUSS seine ersten Versuche zur Elektrizitätslehre an. Auch hier hat er seine wichtigsten Ergebnisse nicht veröffentlicht. Alles hierher Gehörige enthalten die Werke V und Werke XI_1. GAUSS begann damit, daß er über den Magneten des Magnetometers eine Spule schob und es dadurch in ein Galvanometer oder, nach seiner Bezeichnung, in einen « Multiplikator » verwandelte. Die ersten umfangreichen Versuche galten der Bestätigung des Ohmschen Gesetzes und im Zusammenhang damit dem Nachweis der damals noch fraglichen Identität von galvanischer Elektrizität, Thermoelektrizität und Reibungselektrizität. GAUSS prüft auch durch Einschaltung mehrerer Multiplikatoren in seine sehr langen Leitungen nach, ob die Stromstärke an allen Stellen der Leitung die gleiche ist oder ob sich ein Abfall in ihr zeigt. Im Zusammenhang mit der Konstanz der Stromstärke zeigten diese Versuche, daß sich die Fortpflanzungsgeschwindigkeit der Elektrizität in Leitern mit den GAUSS zur Verfügung stehenden Hilfsmitteln nicht messen ließ. Das stand im Einklang mit der Feststellung von WHEATSTONE, daß diese Geschwindigkeit von der Größenordnung der Lichtgeschwindigkeit ist. Bei seinen Untersuchungen fand GAUSS auch die später von KIRCHHOFF wieder entdeckten Gesetze über das Verhalten elektrischer Ströme in Leitungsnetzen und die Herleitung dieser Gesetze aus einem einzigen Minimalprinzip. Die physikalische Deutung dieses Prinzips, daß bei der tatsächlichen Stromverteilung die Joulesche Wärme ein Minimum wird, blieb ihm unbekannt. Nicht eingehen können wir auf Bemerkungen von GAUSS, in denen er mit divinatorischer Sicherheit richtige Erkenntnisse niedergelegt hat, aus denen aber keine Arbeiten entsprungen sind, so z.B. über die Voltasche Spannungsreihe und die Kontaktelektrizität. Hier zieht er zur Erklärung Molekularkräfte heran.

Zur Ausführung der Versuche über das Ohmsche Gesetz hatte W. WEBER die Sternwarte mit dem Physikalischen Institut durch Hin- und Rückleitung verbunden. Schon 1833 kam GAUSS darauf, diese Leitung für eine telegraphische Verbindung mittels galvanischer Elektrizität zu benutzen. Dies erwies sich als unzweckmäßig, da die Zeichengebung zu langsam war. Erst als GAUSS 1835 darauf kam, zum Telegraphieren Induktionsstöße zu benutzen, die er durch Hin- und Herschieben einer Spule auf einem Magneten erhielt (eine solche Einrichtung nennt er « Induktor »), wurde die Nachrichtenübermittlung hinreichend schnell. Verbessert wurde sie noch dadurch, daß GAUSS die Schwingungen des Galvanometers auf der Empfangsseite durch eine Wirbelstromdämpfung rasch abklingen ließ. GAUSS begnügte sich nicht mit einer improvisierten Apparatur, sondern vereinigte alle Teile zu einem zweckmäßig eingerichteten Instrument. So wenig GAUSS vom wissenschaftlichen Standpunkt aus Wert auf seine Erfindung legte, so sehr war er sich ihrer praktischen Bedeutung bewußt. Er nahm Verhandlungen auf, um sie für den Zugsicherungsdienst der damals im Bau befindlichen Eisenbahn Leipzig-Dresden nutzbar zu machen. Näheres ist nicht bekannt, da die Akten vernichtet wurden.

Die größte Leistung von GAUSS in der Elektrizitätslehre ist seine Untersuchung der Induktionserscheinungen. Die erste Beobachtung ist, daß die Schwingungen eines Galvanometers rascher zur Ruhe kommen (ein erhöhtes logarithmisches Dekrement besitzen), wenn der Stromkreis in dem die Galvanometerspule liegt, geschlossen wird. Bei Serienschaltung von mehreren Galvanometern beobachtet GAUSS, daß Schwingungen eines Galvanometers bei den anderen erzwungene Schwingungen mit Phasenverschiebung hervorrufen, die sich den Eigenschwingungen überlagern. Aus dem Briefwechsel mit OLBERS geht hervor, daß GAUSS und WEBER 1835/36 unter Verwendung permanenter Magnete eine magnetelektrische Maschine bauten. Diese benutzten sie zu mannigfachen Versuchen, z.B. zur Zerlegung des Wassers. Über Einzelheiten sind wir nicht genügend orientiert. Diese Maschine erscheint als Fortentwicklung der von GAUSS viel benutzten Induktoren. GAUSS' theoretische Entwicklungen zur Induktion schließen an AMPÈRE an. Dieser hatte die magnetische Wirkung eines geschlossenen Stromes ersetzt durch die einer in den Stromkreis eingespannten magnetischen Doppelschicht. Davon macht GAUSS Gebrauch, um die Wirkungen zweier geschlossener Stromkreise aufeinander bei Schwankungen der Stromstärke in dem einem Stromkreis zu untersuchen. Das Potential der erwähnten Doppelschicht in einem Punkt P ist nichts anderes, als die Stromstärke J mal der scheinbaren Größe oder, wie die Physiker sagen, dem räumlichen Winkel, unter dem die Doppelschicht von P aus erscheint. Befindet sich in P ein positiver magnetischer Einheitspol und läßt man P auf einer geschlossenen Kurve wandern, so ergibt sich aus den früher erwähnten Eigenschaften der scheinbaren Größe, daß sich dabei

bei jedem Durchgang durch die Doppelschicht ein Sprung des Potentials um $\pm 4\pi J$ ergibt, d.h. daß die Arbeit bei dieser Wanderung $A = 4m\pi J$ ist, wenn m die Anzahl der Umschlingungen des Stromkreises durch die von P durchlaufene Kurve ist. So wird Gauss zu dem topologischen Problem geführt, die Anzahl der Umschlingungen zweier geschlossener Kurven zu ermitteln. Die Formel für A enthält in moderne Form umgeschrieben, das Grundgesetz des Elektromagnetismus ($m = 1$). Im Zusammenhang damit tritt bei Gauss auch das Vektorpotential eines linearen Stromes auf.

Die Untersuchungen von Gauss über die Elementargesetze der Kraft zwischen Stromelementen sind für uns nur von historischer Bedeutung. Wohl das wichtigste Ergebnis von Gauss ist folgendes: Ein Element eines Stromes erfährt eine Kraft von dem Magnetfeld eines anderen Stromes. Diese Kraft besitzt ein Potential. Gauss erhält dieses, indem er mit Hilfe des Biot-Savartschen Gesetzes die virtuelle Arbeit bei einer Verrückung des Stromelementes berechnet. Mit der Aufstellung dieses « elektrodynamischen Potentials » und den aus ihm gezogenen physikalischen Folgerungen wurde von Gauss eine der bedeutendsten Leistungen von Franz Neumann um 12 Jahre vorausgenommen.

Die Zusammenarbeit zwischen Gauss und dem auf seine Veranlassung nach Göttingen berufenen W. Weber war so eng, daß sich in vielen Fällen der Anteil beider nicht genau scheiden läßt. Doch war bis 1840 Gauss in dieser Arbeitsgemeinschaft zweifellos der führende Kopf. Diese Zusammenarbeit wurde jäh unterbrochen durch den Staatsstreich des Königs Ernst August von Hannover bei seiner Thronbesteigung 1837, der die geltende Verfassung aufhob und durch ihre reaktionäre Vorgängerin ersetzte. Da Weber zu den sieben protestierenden Göttinger Professoren gehörte, wurde er seines Amtes entsetzt. Zwar blieb er bis zu seiner Berufung nach Leipzig noch in Göttingen, aber dann fanden durch die Trennung die physikalischen Arbeiten von Gauss ein Ende. Webers Rückkehr nach Göttingen 1849 führte zu einer Erneuerung der Zusammenarbeit, aber mit geringerer Wirkung.

Die wichtigen Untersuchungen von Gauss zur geometrischen Optik hängen mit seiner astronomischen Tätigkeit zusammen. Ausgehend von der Dioptrik Eulers berechnet er 1810 für Repsold ein Fernrohrobjektiv, das nach einem ersten Mißerfolg des Schleifers hervorragend gut ausfällt. Wie in diesem Fall regten ihn immer äußere Anlässe zur Beschäftigung mit Fragen der Optik an. So war es bei den dialytischen Fernrohren. Diese wurden konstruiert, weil damals die Herstellung genügend großer einwandfreier Stücke von Flintglas ein technisch schwieriges Problem war. Ein anderer Anlaß war die theoretisch mangelhafte Bestimmung der Brennweite des Königsberger Heliometerobjektivs durch Bessel. Die Rückwirkungen der Wellennatur des Lichtes auf die Ab-

bildung durch optische Instrumente hat GAUSS zu betrachten begonnen. Unter dem Druck anderweitiger Arbeiten brach er diese Untersuchungen ab und kam nie mehr darauf zurück. Daher befriedigen die ganz im Bereich der geometrischen Optik verbleibenden Entwicklungen mit ihrer Beschränkung auf achsennahe Bündel nicht mehr. Dennoch haben sie in der geschichtlichen Entwicklung eine äußerst wichtige Rolle gespielt.

Entsprechend den Gewohnheiten der Zeit berechnet GAUSS die Objektive zunächst so, daß die Bildfeldwölbung für die Randstrahlen und der chromatische Fehler für achsennahe Strahlen beseitigt wird. Man kann nun zu jeder Kronglaslinse einer gegebenen Brennweite eine Flintglaslinse angeben, daß diese Forderungen erfüllt sind. Dabei ist das Verhältnis der Krümmungsradien der Kronglaslinse noch willkürlich. Das irrige Ergebnis eines anderen Forschers, daß man diese Tatsache nicht zu einer weiteren Verbesserung der Objektive ausnutzen könne, veranlaßte GAUSS zur Nachprüfung. In der Arbeit « Über die achromatischen Doppelobjektive, besonders in Rücksicht der vollkommenern Aufhebung der Farbenzerstreuung » (Werke V) von 1817 zeigt er, daß bei geeigneter Wahl der Krümmungsradien der Kronglaslinse der chromatische Fehler auch für die Randstrahlen beseitigt werden kann. Die erfolgreiche Herstellung eines solchen Objektivs gelang erst 1860 STEINHEIL, der darlegt, daß die früheren Mißerfolge teilweise in technischen Schwierigkeiten, teilweise in einem ungenügenden Verständnis der Entwicklungen von GAUSS ihren Grund hatten. GAUSS überlegt auch, ob es zweckmäßig sei, die möglichen Korrekturen gerade für die Zentral- und die Randstrahlen durchzuführen und bejaht das zunächst, gibt aber später eine Richtigstellung.

Von 1840-1841 beschäftigt sich GAUSS mit den dialytischen Fernrohren. Bei diesen war die Flintglaslinse von der Kronglaslinse getrennt und an eine Stelle verlegt, an der das Lichtstrahlenbündel sehr eng ist. Die Trennung von Kron- und Flintglaslinse nötigt dann dazu, diese letztere durch eine Kronglaslinse zu ergänzen. GAUSS bemerkt sofort, daß ein solches Objektiv nicht achromatisch sein könne. Er findet aber bald, daß dies ganz unwesentlich ist, sondern daß es darauf ankommt, daß das ganze, aus Objektiv und Okular (und u.U. auch dem Auge) bestehende optische System achromatisch ist.

Die wichtigste optische Arbeit von GAUSS sind die « Dioptrischen Untersuchungen » (Werke V) von 1840. Zu diesen sei vorweg bemerkt, daß die Einführung der Hauptebenen zur Erkenntnis führt, daß Gegenstandsraum und Bildraum kollinear auf einander bezogen sind. Es ist sehr merkwürdig, daß dies bei GAUSS nirgends gesagt wird, trotz des hieraus folgenden Gewinns an Einfachheit und Klarheit. GAUSS verfolgt den Gang eines Lichtstrahls durch ein zentriertes System brechender Kugelflächen unter Beschränkung auf achsennahe Strahlen. Die rechnerische Verfolgung des Lichtstrahls durch das System ist eine elementare Aufgabe der analytischen Geometrie. Aus ihr kann man

entnehmen, wo das Bild P^* eines auf dem einfallenden Strahl liegenden Punktes P liegt. Das Verhältnis des Abstandes von P^* von der Achse zum Abstand von P von der Achse nennt GAUSS Lateralvergrößerung. Der Ort der Punkte P, welche Bildpunkte P^* mit der Lateralvergrößerung 1 ergeben, ist eine zur Achse senkrechte Ebene, das gleiche gilt für die P^*. Diese Ebenen sind die beiden Hauptebenen, ihre Schnitte mit der Achse die Hauptpunkte. Verlängert man einen einfallenden Strahl (ohne Brechung!) bis zu seinem Schnitt Q mit der ersten Hauptebene, so geht der ausfallende Strahl rückwärts verlängert durch Q^*. Daraus folgt, daß man das brechende System ersetzen kann durch eine einzige brechende Kugelfläche von bekanntem Radius, wenn die Brechzahlen des ersten und letzten Mediums verschieden sind, und durch eine einzige unendlich dünne Linse bekannter Brennweite, wenn die Brechzahlen gleich sind. Es folgt die Erklärung der Brennpunkte und der Brennweiten als Abstände der Brennpunkte von den zugehörigen Hauptpunkten. Besonders behandelt wird der singuläre Fall, daß beide Brennweiten unendlich sind, wie es bei dem auf Unendlich eingestellten Fernrohr der Fall ist.

* * *

Bei unserer Darstellung der wissenschaftlichen Leistungen von GAUSS mußte eine Unzahl wichtiger und bedeutender Leistungen von GAUSS in der Mathematik und Physik übergangen werden, so z.B. seine Beschäftigung mit der Versicherungsmathematik. Jedoch werden die besprochenen Arbeiten wohl ein ausreichendes Bild von der einzigartigen wissenschaftlichen Größe des Mannes ergeben.

LITERATURVERZEICHNIS

[1] CARL FRIEDRICH GAUSS: *Werke*, herausgegeben von der Gesellschaft der Wissenschaften zu Göttingen, 12 Bände, 1863-1929.

[2] Zum Teil im Vorabdruck erschienen unter dem Titel: *Materialien zu einer wissenschaftlichen Biographie von Gauss.* Gesammelt von F. KLEIN, M. BRENDEL und L. SCHLESINGER. 8 Hefte, Leipzig 1911-1920.

[3] a) *Briefwechsel zwischen Gauss und Bessel* (herausg. von A. AUWERS), Leipzig, 1880.
b) *Briefwechsel zwischen C. F. Gauss und W. Bolyai* (herausg. von F. SCHMIDT und P. STÄCKEL), Leipzig, 1899.
c) *Briefwechsel zwischen C. F. Gauss und Chr. L. Gerling* (herausg. von CLEMENS SCHAEFER), Berlin, 1927.
d) *Briefwechsel zwischen C. F. Gauss und P. G. Lejeune Dirichlet*, enthalten in: DIRICHLET: *Werke* II (herausg. von L. KRONECKER), Berlin, 1897.
e) *Briefe zwischen A. v. Humboldt und Gauss* (herausg. von K. BRUHNS), Leipzig, 1877.

f) *Briefe von C. F. Gauss an B. Nicolai* (herausg. von W. Valentiner), Karlsruhe, 1877.

g) *Wilhelm Olbers, sein Leben und seine Werke* (herausg. von C. Schilling), Band II (Briefwechsel zwischen Olbers und Gauss), Berlin, 1900-1909.

h) *Briefwechsel zwischen C. F. Gauss und H. C. Schumacher* (herausg. von C. A. F. Peters), Band I-VI, Altona, 1860-1865.

[4] F. Klein: *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*, Teil I, bearbeitet von R. Courant und O. Neugebauer, Berlin, 1936.

[5] C. F. Gauss: *Bestimmung der Anziehung eines elliptischen Ringes. – Nachlaß zur Theorie des arithmetisch-geometrischen Mittels und der Modulfunktion* (Übersetzt und herausg. von H. Geppert), in *Ostwalds Klassiker der exakten Wissenschaften*, 225, Leipzig, 1927.

SUPPLEMENTO AL VOLUME II, SERIE X
DEL NUOVO CIMENTO

N. 5, 1955
2° Semestre

ERRATA-CORRIGE

G. BERNARDINI: **Lectures on Photoproduction.** *Supplemento al Vol. II, Serie X del Nuovo Cimento,* N. 1, 1955, 2° Semestre, pag. 104-138.

Because of a combination of circumstances and mainly because of the author's fault the galleys of this article have been revised only by the Editor's reviser. He did an excellent job and the author wants to express his appreciation to him. However few confusing misprints remained mostly in tables and formulae. Some of them are listed below.

Pag. 104,	*title*:	*to be added to Author's name*:	« Now Professor at the University of Illinois, Urbana, Illinois ».			
» 111,	*line 7th from bottom*	:	*instead of*	is easy	*replace*	is not easy (see later)
» 114,	*Table I, column 6th*	:	» »	387	»	340
» 115,	*line 13th from top*	:	» »	A_1	»	A_2
» 117,	*Table II, title*	:	» »	10^{-13}	»	10^{-30}
» 119,	*formulae* (6), (6′)	:	*delete the coefficient*	$\frac{2}{\mu^2}$		
» 120,	*line 9th from top*	:	*instead of*	Appendix II	*replace*	Section **3**
» 120,	*lines 18th–19th from top*	:	» »	Appendix II	»	Section **3**
» 125,	*formulae* (7), *line 2nd*	:	» »	${}_{-1}Y\chi_{l,2}$	»	$\chi_{-1}Y_{l,2}$
	» (9), » *1st*	:	» »	, $+1$;	»	, ± 1;
» 130,	*last line*	:	» »	$P\rightleftarrows\vert$	»	$\vert p\rightleftarrows$
» 132,	*table, head of the column 4th*:		» »	$C(3/2,$	»	$C_{1,\frac{1}{2}}(3/2,$
» 133,	*table, head of the column 4th*:		» »	$C(3/2,$	»	$C_{2,\frac{1}{2}}(3/2,$
	Table II, column 4th	*line 3th*:	» »	no	»	yes + no
		» 4th:	» »	no	»	yes + no
		» 5th:	» »	yes	»	yes + no
		» 6th:	» »	yes	»	yes + no
» 134,	line *15th from top*	:	» »	matrix	»	S-matrix
» 135,	» 6th » »	:	» »	$\mathrm{d}\sigma/\mathrm{d}\Omega$	»	flux $\times\ \mathrm{d}\sigma/\mathrm{d}\Omega$
» 136,	*last line*	:	» »	χ_{+1}	»	χ_{-1}
» 137,	*3rd* » *from top*	:	» »	χ_{-2}	»	χ_{-1}
	9th » » *bottom*	:	» »	$\overline{b}_k$	»	b_k
	4th » » »	:	» »	$\times$	»	$+$

G. BERNARDINI

INDICE DEL SUPPLEMENTO

AL VOLUME II, SERIE X, DEL

NUOVO CIMENTO

Anno 1955

Rendiconti del Corso tenuto nella Villa Monastero a Varenna dal 18 Luglio al 7 Agosto 1954 . . . . . pag. 1

Commemorazione di ENRICO FERMI tenutasi a Varenna e a Como il 6 Agosto 1955 . . . . . » 471

Onoranze a MACEDONIO MELLONI nel 1° Centenario della sua morte: Rendiconti del XL Congresso Nazionale di Fisica e del Convegno Internazionale di Studi sull'Infrarosso - Parma 3-7 Settembre 1954 . . . . . » 491

B. JOUVET – A Realistic Theory of Mesons . . . . . » 941

M. SIMONETTA – Cinetica e meccanismo delle reazioni di sostituzione nucleofila » 969

N. WIENER and A. SIEGEL – The Differential-Space Theory of Quantum Systems . . . . . » 982

D. BOHM and W. SCHÜTZER – The General Statistical Problem in Physics and the Theory of Probability . . . . . » 1004

D. BRINI, L. PELI, O. RIMONDI and P. VERONESI – On Large Scintillation Counters. . . . . » 1048

G. LÜDERS – Über den Einfluß von Fehlern des magnetischen Feldes auf die Betatronschwingungen im Synchrotron mit starker Stabilisierung . . . » 1075

K. URBANIK – Some Remarks on the Asymptotic Behaviour of the Cosmic Ray Cascade for Large Depth of the Absorber. - I. Estimation of the Factorial Moments . . . . . » 1147

J. ŁOPUSZAŃSKI – Some Remarks on the Asymptotic Behaviour of the Cosmic Ray Cascade for Large Depth of the Absorber. - II. Asymptotic Behaviour of the Probability Distribution Function . . . . . » 1150

J. ŁOPUSZAŃSKI – Some Remarks on the Asymptotic Behaviour of the Cosmic Ray Cascade for Large Depth of the Absorber. - III. Evaluation of the Distribution Function. . . . . » 1161

R. FIESCHI – Thermodynamical Theory of Galvanomagnetic and Thermomagnetic Phenomena . . . . . » 1168

T. BOLELLI – Lingua e stile di Galileo . . . . . » 1173

C. LANCZOS – Albert Einstein and the Theory of Relativity . . . . . » 1193

M. KRAFFT – Carl Friedrich Gauß . . . . . » 1221

G. BERNARDINI – Errata-Corrige . . . . . » 1256

Fine del *Supplemento* al Volume II, Serie X,
del *Nuovo Cimento*, 1955

PROPRIETÀ LETTERARIA RISERVATA

Direttore responsabile: G. POLVANI

Tipografia Compositori - Bologna

Questo fascicolo del *Supplemento* è stato licenziato dai torchi il 24-IV-1956